On the Spectral and Energy Efficiencies of Full-Duplex Cell-Free Massive MIMO

Hieu V. Nguyen, *Student Member, IEEE*, Van-Dinh Nguyen, *Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE*, Shree Krishna Sharma, *Senior Member, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Oh-Soon Shin, *Member, IEEE*

***Abstract*—In-band full-duplex (FD) operation is practically more suited for short-range communications such as WiFi and small-cell networks, due to its current practical limitations on the self-interference cancellation. In addition, cell-free massive multiple-input multiple-output (CF-mMIMO) is a new and scalable version of MIMO networks, which is designed to bring service antennas closer to end user equipments (UEs). To achieve higher spectral and energy efficiencies (SE-EE) of a wireless network, it is of practical interest to incorporate FD capability into CF-mMIMO systems to utilize their combined benefits. We formulate a novel and comprehensive optimization problem for the maximization of SE and EE in which power control, access point-UE (AP-UE) association and AP selection are jointly optimized under a realistic power consumption model, resulting in a difficult class of mixed-integer nonconvex programming. To tackle the binary nature of the formulated problem, we propose an efficient approach by exploiting a strong coupling between binary and continuous variables, leading to a more tractable problem. In this regard, two low-complexity transmission designs based on zero-forcing (ZF) are proposed. Combining tools from inner approximation framework and Dinkelbach method, we develop simple iterative algorithms with polynomial computational complexity in each iteration and strong theoretical performance guaranteed. Furthermore, towards a robust design for FD CF-mMIMO, a novel heap-based pilot assignment algorithm is proposed to mitigate effects of pilot contamination. Numerical results show that our proposed designs with realistic parameters significantly outperform the well-known approaches (i.e., small-cell and collocated mMIMO) in terms of the SE and EE. Notably, the proposed ZF designs require much less execution time than the simple maximum ratio transmission/combining.**

***Index Terms*—Cell-free massive multiple-input multiple-output, energy efficiency, full-duplex radio, inner approximation, spectral efficiency, successive interference cancellation.**

This work was supported in part by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-00724, Development of Beyond 5G Mobile Communication Technologies (Ultra-Reliable, Low-Latency, and Massive Connectivity) and Combined Access Technologies for Cellular-based Industrial Automation Systems), in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1A2C1084834, No. 2017R1A5A1015596), in part by the ERC project AGNOSTIC and FNR ECLECTIC, and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), through its Discovery program. Part of this work will be presented at the IEEE International Conference on Communications 2020, Dublin, Ireland [1]. *(Corresponding author: Oh-Soon Shin.)*

H. V. Nguyen and O.-S. Shin are with the School of Electronic Engineering & Department of ICMC Convergence Technology, Soongsil University, Seoul 06978, South Korea (e-mail: {hieuvnguyen, osshin}@ssu.ac.kr).

V.-D. Nguyen is with the Interdisciplinary Centre for Security, Reliability and Trust (SnT) – University of Luxembourg, L-1855 Luxembourg. He was with the Department of ICMC Convergence Technology, Soongsil University, Seoul 06978, South Korea (email: dinh.nguyen@uni.lu).

O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1X3C5, Canada (e-mail: odobre@mun.ca).

S. K. Sharma, S. Chatzinotas, and B. Ottersten are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT) – University of Luxembourg, L-1855 Luxembourg (e-mail: {shree.sharma, symeon.chatzinotas, bjorn.ottersten}@uni.lu)

I. Introduction

Peak data rates in the order of tens of Gbits/s, massive connectivity and seamless area coverage requirement along with different use cases are expected in beyond 5G networks [2]–[4]. Multiple-antenna technologies, which offer extra degrees-of-freedom (DoF) have been a key element to provide huge spectral efficiency gains of modern wireless communication systems [5], [6]. However, multiple-antenna systems based on half-duplex (HD) radio will apparently reach their capacity limits in near future due to under-utilization of radio resources. In addition, the use of multiple-antenna also causes a serious concern over the global climate and tremendous electrical costs due to the number of associated radio frequency (RF) elements [7]. Consequently, spectral efficiency (SE) and energy efficiency (EE) will certainly be considered as major figure-of-merit in the design of beyond 5G networks.

In-band full-duplex (FD) has been envisaged as a key enabling technology to improve the SE of traditional wireless communication systems [8], [9]. By enabling downlink (DL) and uplink (UL) transmissions on the same time-frequency resource, FD radios are expected to increase the SE of a wireless link over its HD counterparts by a factor close to two [10], [11]. The main barrier in implementing FD is the self-interference (SI) that leaks from the transmitter to its own receiver on the same device. Fortunately, recent advances in active and passive SI suppression techniques have been successful to bring the SI power at the background noise level [12], [13], thereby making FD a realistic technology for modern wireless systems. However, there still exists a small, but not negligible, amount of SI due to imperfect SI suppression, referred to as *residual* SI (RSI). As a result, FD-enabled base station (BS) systems have been widely studied in small-cell (SC) cellular networks [14]–[22], where the residual SI can be further handled by power control algorithms.

Recently, a new concept of multiple-input multiple-output (MIMO) networks and distributed antenna systems (DAS), called cell-free massive MIMO (CF-mMIMO), has been proposed to overcome the inter-cell interference, as well as to provide handover-free and balanced quality-of-experience (QoE) services for cell-edge user equipments (UEs) [5], [23]–[28]. In CF-mMIMO, a very large number of access point (AP) antennas are distributed over a wide area to coherently serve numerous UEs in the same resources; this inherits key characteristics of collocated massive MIMO (Co-mMIMO) networks, such as favorable propagation and channel hardening [29], [30]. CF-mMIMO has significantly better performances in

terms of SE and EE, compared to small-cell [23] and Co-mMIMO networks [28], respectively. It can be easily foreseen that performance gains of CF-mMIMO come from the joint processing of a large number of distributed APs at a central processing unit (CPU). The CPU is essentially the same as the edge-cloud computing in cloud radio access networks (C-RANs). Thus, C-RAN can be viewed as an enabler of CF-mMIMO [5].

A. Motivation

From the aforementioned reasons, it is not too far-fetched to envisage a wireless system employing the FD technology in CF-mMIMO, called FD CF-mMIMO. It is expected to reap all key advantages of FD and CF-mMIMO, towards enhancing the SE and EE performances of future wireless networks. More importantly, FD CF-mMIMO can be considered as a practical and promising technology for beyond 5G networks since low-power and low-cost FD-enabled APs are well suited for short-range transmissions between APs and UEs. Despite the clear benefits of these two technologies, FD CF-mMIMO poses the following obvious challenges on radio resource allocation problems: (i) Residual SI still remains a challenging task in the design of FD CF-mMIMO, having a negative impact on its potential performance gains; (ii) A large number of APs and legacy UEs result in stronger inter-AP interference (IAI) and co-channel interference (CCI, caused by the UL transmission to DL UEs), compared to traditional FD cellular networks [14]–[22]; (iii) FD CF-mMIMO increases the network power consumption due to additional number of APs. It has been noted that low power APs consume about 30% of the total power consumption of a mobile network operator [31]. These motivate us to investigate a joint design of precoder/receiver, AP-UE association and AP selection along with an efficient transmission strategy to attain the optimal SE and EE performances of FD CF-mMIMO systems.

B. Review of Related Literature

FD small-cell systems have been investigated in many prior works. For example, the authors in [15] studied a single-cell network with the aim of maximizing the SE under the assumption of perfect channel state information (CSI). This work was generalized in [16] where user grouping and time allocation were jointly designed. To accelerate the use of FD operation, [17] proposed a half-array antenna mode selection to mitigate the effect of residual SI and CCI by serving UEs in two separate phases. This design is capable of enabling hybrid modes of HD and FD to utilize a full-array antenna at the BS. The application of FD to emerging subjects has also been investigated, including FD non-orthogonal multiple access [18], FD physical layer security [19] and FD wireless-powered MIMO [20], [32]. The SE maximization for FD multi-cell networks was considered in [21] and with the worst-case robust design in [22], where coordinated multi-point transmission was adopted. It is widely believed that this interference-limited technique can no longer provide a high edge throughput and requires a large amount of backhaul signaling to be shared among BSs. In addition, the common transmission design used in these works is linear beamforming for DL and minimum mean square error and successive interference cancellation (MMSE-SIC) receiver for UL. Although such a design can provide a very good performance, it is only suitable for networks of small-to-medium sizes.

CF-mMIMO has recently received considerable attention. In particular, the work in [23] first derived closed-form expressions of DL and UL achievable rates which confirm the SE gain of the CF-mMIMO over a small-cell system. Assuming mutually orthogonal pilot sequences assigned to UEs, [24] analyzed impacts of the power allocation for DL transmission using maximum ratio transmission (MRT) and zero-forcing (ZF). The results showed that the achievable per-user rates of CF-mMIMO can be substantially improved, compared to those of small-cell systems. To further improve the network performance of CF-mMIMO, a beamformed DL training was proposed in [25]. The authors in [26] examined the problem of maximizing the minimum signal-to-interference-plus-noise ratio (SINR) of UL UEs subject to power constraints. More recently, the EE problem for DL CF-mMIMO was investigated using ZF in [27] and MRT in [28], by taking into account the effects of power control, non-orthogonality of pilot sequences, channel estimation and hardware power consumption. Two simple AP selection schemes were also proposed in [28] to reduce the backhaul power consumption. The conclusion in these works was that ZF and MRT beamforming methods can offer excellent performance when the number of APs is large.

Despite its potential, there is only a few attempts on characterizing the performance of FD CF-mMIMO in the literature. In this regard, the authors in [33] analyzed the performance of FD CF-mMIMO with the channel estimation taken into account, where all APs operate in FD mode. By a simple conjugate beamforming/matched filtering transmission design, it was shown that such a design requires a deep SI suppression to unveil the performance gains of the FD CF-mMIMO system over HD CF-mMIMO one. Tackling the imperfect CSI and spatial correlation was studied in [34], showing that FD CF-mMIMO with a genetic algorithm-based user scheduling strategy is able to help alleviating the CCI and obtaining a significant SE improvement. However, the effects of SI, IAI and CCI are not fully addressed in the above-cited works, leading to the need of optimal solutions, which is the focus of this paper.

C. Research Gap and Main Contributions

Though in-depth results of multiple-antenna techniques were presented for HD [5] and FD operations [14]–[22], they are not very practical for FD CF-mMIMO due to the large size of optimization variables. In addition, a direct application of CF-mMIMO in [23]–[28] to FD CF-mMIMO systems would result in a poor performance since the additional interference (i.e., residual SI, IAI and CCI) is strongly involved in both DL and UL transmissions. In FD CF-mMIMO systems, the number of APs is large, but still finite, and thus the effects of SI, IAI and CCI are acute and unavoidable. These issues have not been fully addressed in [33], [34]. Meanwhile, the EE performance in terms of bits/Joule is considered as a key

performance indicator for green communications [31], which is even more important in FD CF-mMIMO systems due to a large number of deployed APs. To the authors' best knowledge, the EE maximization, that takes into account the effects of imperfect CSI, power allocation, AP-DL UE association, AP selection, and load-dependent power consumption, has not been previously studied for FD CF-mMIMO.

In the above context, this paper considers an FD CF-mMIMO system under time-division duplex (TDD) operation, where FD-enabled multiple-antenna APs simultaneously serve many UL and DL UEs on the same time-frequency resources. The network SE and EE of the FD CF-mMIMO system are investigated under a relatively realistic power consumption model. Furthermore, we propose an efficient transmission design for FD CF-mMIMO to resolve practical restrictions given above. More precisely, power control and AP-DL UE association are jointly optimized to reduce network interference. Inspired by the ZF method [35], two simple, but efficient transmission schemes are proposed. It is well-known that in massive MIMO networks, ZF-based schemes achieve close performance to the MMSE ones but with much less complexity [30]. In contrast with [33] and [34], AP selection is also taken into account to preserve the hardware power consumption. We note that the EE objective function strikes the balance between the SE and total power consumption. The main contributions of this paper are summarized as follows:

- Aiming at the optimization of SE and EE, we introduce new binary variables to establish AP-DL UE association and AP selection. This design not only helps mitigate network interference (residual SI, IAI and CCI) but also saves power consumption (i.e., some APs can be switched off if necessary). In our system design, APs can be automatically switched between FD and HD operations, which allows to exploit the full potential of FD CF-mMIMO.
- We formulate a generalized maximization problem for SE and EE by incorporating various aspects such as joint power control, AP-DL UE association and AP selection, which belongs to the difficult class of mixed-integer nonconvex optimization problem. To develop an unified approach to all considered transmission schemes, we first provide fundamental insights into the structure of the optimal solutions of binary variables, and then transform the original problem into a more tractable form.
- Two low-complexity transmission schemes for DL/UL are proposed, *namely* ZF and improved ZF (IZF) by employing principal component analysis (PCA) for DL and SIC for UL. The latter based on the orthonormal basis (ONB) is capable of canceling multiuser interference (MUI) and further alleviating residual SI and IAI. We then employ the combination of inner convex approximation (ICA) framework [36], [37] and Dinkelbach method [38] to develop iterative algorithms of low-computational complexity, which converge rapidly to the optimal solutions as well as require much lower execution time than the simple maximum ratio transmission/combining (MRT/MRC).
- Towards practical applications, we further consider a robust design under channel uncertainty, where the UL training is taken into account. To this end, we develop a novel heap-based pilot assignment algorithm to reduce both the pilot contamination and training complexity.
- Extensive numerical results confirm that the proposed algorithms greatly improve the SE and EE performance over the current state-of-the-art approaches, i.e., SC-MIMO and Co-mMIMO under both HD and FD operation modes. It also reveals the effectiveness of joint AP selections in terms of the achieved EE performance.

D. Paper Organization and Notation

The remainder of this paper is organized as follows. The FD CF-mMIMO system model is introduced in Section II. The formulation of the optimization problem and the derivation of its tractable form are provided in Section III. Two ZF-based transmission designs along the proposed algorithm are presented in Sections IV and V. The pilot assignment algorithm is presented in Section VI. Numerical results are given in Section VII, while Section VIII concludes the paper.

Notation: Bold lower and upper case letters denote vectors and matrices, respectively. $\mathbf{X}^T$ and $\mathbf{X}^H$ represent normal transpose and Hermitian transpose of $\mathbf{X}$, respectively. $\mathtt{Tr}(\cdot)$, $\|\cdot\|$ and $|\cdot|$ are the trace, Euclidean norm and absolute value, respectively. $\mathbf{a} \preceq \mathbf{b}$ stands for the element-wise comparison of vectors. $\mathtt{diag}(\mathbf{a})$ returns the diagonal matrix with the main diagonal constructed from elements of $\mathbf{a}$. $\mathbb{C}$ and $\mathbb{R}$ denote the space of complex and real matrices, respectively. Finally, $x \sim \mathcal{CN}(0, \sigma^2)$ and $x \sim \mathcal{N}(0, \sigma^2)$ represent circularly symmetric complex and real-valued Gaussian random variable with zero mean and variance σ^2, respectively.

II. System Model

A. Transmission Model

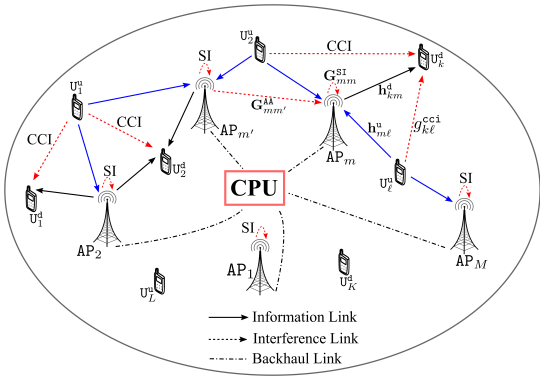


Fig. 1. An illustration of the full-duplex cell-free massive MIMO system.

An FD CF-mMIMO system operated in TDD mode is considered, where the set $\mathcal{M} \triangleq \{1, 2, \cdots, M\}$ of $M = |\mathcal{M}|$ FD-enabled APs simultaneously serves the sets $\mathcal{K} \triangleq \{1, 2, \cdots, K\}$ of $K = |\mathcal{K}|$ DL UEs and $\mathcal{L} \triangleq \{1, 2, \cdots, L\}$ of $L = |\mathcal{L}|$ UL UEs in the same time-frequency resource, as illustrated in Fig. 1. The total number of APs' antennas is $N = \sum_{m \in \mathcal{M}} N_m$ where N_m is the number of antennas at AP m, while each UE has a single-antenna. We assume that APs, DL and UL UEs are randomly placed in a wide area. All APs

are equipped with FD capability by circulator-based FD radio prototypes [12], [13], which are connected to the CPU through perfect backhaul links with sufficiently large capacities (i.e., high-speed optical ones) [23]. In this paper, we focus on slowly time-varying channels, and thus, conveying the CSI via the backhaul links occurs less frequently than data transmission.

We assume that data transmission is performed within a coherence interval, which is similar to TDD operation in the context of massive MIMO [30]. Based on the joint processing at the CPU, the message sent by an UL UE is decoded by aggregating the received signals from all active APs due to the UL broadcast transmission. In addition, APs are geographically distributed in a large area, and thus, each DL UE should be served by a subset of active APs with good channel conditions [23], [28]. This is done by introducing new binary variables to establish AP-DL UE associations. Such a design offers the following two obvious advantages: (i) improving the SE for a given system bandwidth and power budget of APs, while still ensuring the quality of service (QoS) for all UEs; (ii) managing the network interference more effectively.

For notational convenience, let us denote the m-th AP, k-th DL UE and ℓ-th UL UE by $\mathtt{AP}_m$, $\mathtt{U}_k^{\mathtt{d}}$ and $\mathtt{U}_\ell^{\mathtt{u}}$, respectively. The channel vectors and matrices from $\mathtt{AP}_m \rightarrow \mathtt{U}_k^{\mathtt{d}}$, $\mathtt{U}_\ell^{\mathtt{u}} \rightarrow \mathtt{AP}_m$, $\mathtt{U}_\ell^{\mathtt{u}} \rightarrow \mathtt{U}_k^{\mathtt{d}}$ and $\mathtt{AP}_{m'} \rightarrow \mathtt{AP}_m, \forall m' \in \mathcal{M}$ are denoted by $\mathbf{h}_{km}^{\mathtt{d}} \in \mathbb{C}^{1\times N_m}$, $\mathbf{h}_{m\ell}^{\mathtt{u}} \in \mathbb{C}^{N_m\times 1}$, $g_{k\ell}^{\mathtt{cci}} \in \mathbb{C}$ and $\mathbf{G}_{mm'}^{\mathtt{AA}} \in \mathbb{C}^{N_m\times N_{m'}}$, respectively. Note that $\mathbf{G}_{mm}^{\mathtt{AA}}$ is the SI channel at $\mathtt{AP}_m$, while $\mathbf{G}_{mm'}^{\mathtt{AA}}, \forall m \neq m'$ is referred to as the inter-AP interference (IAI) channel since UL signals received at $\mathtt{AP}_m$ are corrupted by DL signals sent from $\mathtt{AP}_{m'}$. The reason for this is that the SI signal can only be suppressed at the local APs [12], [13]. To differentiate the residual SI and IAI channels, we model $\mathbf{G}_{mm'}^{\mathtt{AA}}$ as follows:

$$\mathbf{G}_{mm'}^{\mathtt{AA}} = \begin{cases} \sqrt{\rho_{mm}^{\mathtt{RSI}}}\mathbf{G}_{mm}^{\mathtt{SI}}, & \text{if } m = m', \\ \mathbf{G}_{mm'}^{\mathtt{AA}}, & \text{otherwise,} \end{cases}$$

where $\mathbf{G}_{mm}^{\mathtt{SI}}$ denotes the fading loop channel at $\mathtt{AP}_m$ which interferes the UL reception due to the concurrent DL transmission, and $\rho_{mm}^{\mathtt{RSI}} \in [0,1)$ is the residual SI suppression (SiS) level after all real-time cancellations in analog-digital domains [10], [17], [19], [20]. The fading loop channel $\mathbf{G}_{mm}^{\mathtt{SI}}$ can be characterized as the Rician probability distribution with a small Rician factor [39], while other channels are generally modeled as $\mathbf{h} = \sqrt{\beta}\ddot{\mathbf{h}}$ with $\mathbf{h} \in \{\mathbf{G}_{mm'}^{\mathtt{AA}}, \mathbf{h}_{km}^{\mathtt{d}}, \mathbf{h}_{m\ell}^{\mathtt{u}}, g_{k\ell}^{\mathtt{cci}}\}$, accounting for the effects of large-scale fading β (i.e., path loss and shadowing) and small-scale fading $\ddot{\mathbf{h}}$ whose elements follow independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ random variables (RVs).

1) Downlink Data Transmission

Let us denote by $x_k^{\mathtt{d}}$ and $x_\ell^{\mathtt{u}}$ the data symbols with unit power (i.e., $\mathbb{E}\big[|x_k^{\mathtt{d}}|^2\big] = 1$ and $\mathbb{E}\big[|x_\ell^{\mathtt{u}}|^2\big] = 1$) intended for $\mathtt{U}_k^{\mathtt{d}}$ and sent from $\mathtt{U}_\ell^{\mathtt{u}}$, respectively. The beamforming vector $\mathbf{w}_{km} \in \mathbb{C}^{N_m\times 1}$ is employed to precode the data symbol $x_k^{\mathtt{d}}$ of $\mathtt{U}_k^{\mathtt{d}}$ in the DL, while p_ℓ denotes the transmit power of $\mathtt{U}_\ell^{\mathtt{u}}$ in the UL. After performing a joint radio resource management algorithm at the CPU, the data of $\mathtt{U}_k^{\mathtt{d}}$ is routed to $\mathtt{AP}_m$ via the m-th backhaul link only if $\|\mathbf{w}_{km}\| > 0$. To do so, let us introduce new binary variables $\alpha_{km} \in \{0,1\}, \forall k \in \mathcal{K}, m \in \mathcal{M}$ to represent the association relationship between $\mathtt{AP}_m$ and $\mathtt{U}_k^{\mathtt{d}}$, i.e., $\alpha_{km} = 1$ implying that $\mathtt{U}_k^{\mathtt{d}}$ is served by $\mathtt{AP}_m$ and $\alpha_{km} = 0$, otherwise. Using these notations, the signal received at $\mathtt{U}_k^{\mathtt{d}}$ can be expressed as

$$y_k^{\mathtt{d}} = \sum\nolimits_{m\in\mathcal{M}} \alpha_{km}\mathbf{h}_{km}^{\mathtt{d}}\mathbf{w}_{km}x_k^{\mathtt{d}} + \underbrace{\sum\nolimits_{\ell\in\mathcal{L}} \sqrt{p_\ell}g_{k\ell}^{\mathtt{cci}}x_\ell^{\mathtt{u}}}_{\text{CCI}} + \underbrace{\sum\nolimits_{m\in\mathcal{M}}\sum\nolimits_{k'\in\mathcal{K}\setminus\{k\}} \alpha_{k'm}\mathbf{h}_{km}^{\mathtt{d}}\mathbf{w}_{k'm}x_{k'}^{\mathtt{d}}}_{\text{MUI}} + n_k, \quad (1)$$

where $n_k \sim \mathcal{CN}(0,\sigma_k^2)$ is the additive white Gaussian noise (AWGN), and σ_k^2 is the noise variance. By treating MUI and CCI as noise, the received SINR at $\mathtt{U}_k^{\mathtt{d}}$ is given as

$$\gamma_k^{\mathtt{d}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha}) = \frac{\sum_{m\in\mathcal{M}}\alpha_{km}|\mathbf{h}_{km}^{\mathtt{d}}\mathbf{w}_{km}|^2}{\chi_k(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})}, \quad (2)$$

where $\chi_k(\mathbf{w},\mathbf{p},\boldsymbol{\alpha}) \triangleq \sum_{m\in\mathcal{M}}\sum_{k'\in\mathcal{K}\setminus\{k\}}\alpha_{k'm}|\mathbf{h}_{km}^{\mathtt{d}}\mathbf{w}_{k'm}|^2 + \sum_{\ell\in\mathcal{L}}p_\ell|g_{k\ell}^{\mathtt{cci}}|^2 + \sigma_k^2$, $\mathbf{w} \triangleq [\mathbf{w}_1^H,\cdots,\mathbf{w}_K^H]^H \in \mathbb{C}^{NK\times 1}$ with $\mathbf{w}_k \triangleq [\mathbf{w}_{k1}^H,\cdots,\mathbf{w}_{kM}^H]^H \in \mathbb{C}^{N\times 1}$, $\mathbf{p} \triangleq [p_1,\cdots,p_L]^T \in \mathbb{R}^{L\times 1}$, and $\boldsymbol{\alpha} \triangleq \{\alpha_{km}\}_{\forall k\in\mathcal{K},m\in\mathcal{M}}$. We note that in (2), α_{km} is equal to α_{km}^2 for any $\alpha_{km} \in \{0,1\}$.

2) Uplink Data Transmission

The received signal at $\mathtt{AP}_m$ can be expressed as

$$\mathbf{y}_m^{\mathtt{u}} = \sum\nolimits_{\ell\in\mathcal{L}}\sqrt{p_\ell}\mathbf{h}_{m\ell}^{\mathtt{u}}x_\ell^{\mathtt{u}} + \underbrace{\sum\nolimits_{m'\in\mathcal{M}}\sum\nolimits_{k\in\mathcal{K}}\alpha_{km'}\mathbf{G}_{mm'}^{\mathtt{AA}}\mathbf{w}_{km'}x_k^{\mathtt{d}}}_{\text{RSI + IAI}} + \mathbf{n}_m, \quad (3)$$

where $\mathbf{n}_m \sim \mathcal{CN}(0,\sigma_{\mathtt{AP}}^2\mathbf{I})$ is the AWGN. The CPU aggregates the received signals from all APs, and the $\mathtt{U}_\ell^{\mathtt{u}}$'s signal is then extracted by using a specific receiver. In general, let us denote the receiver vector to decode the $\mathtt{U}_\ell^{\mathtt{u}}$'s message received at $\mathtt{AP}_m$ by $\mathbf{a}_{m\ell} \in \mathbb{C}^{1\times N_m}$, and thus, the received signal of $\mathtt{U}_\ell^{\mathtt{u}}$ at $\mathtt{AP}_m$ can be expressed as $r_{m\ell}^{\mathtt{u}} = \mathbf{a}_{m\ell}\mathbf{y}_m^{\mathtt{u}}$. Consequently, the post-detection signal for decoding the $\mathtt{U}_\ell^{\mathtt{u}}$'s signal is $r_\ell^{\mathtt{u}} = \sum_{m\in\mathcal{M}} r_{m\ell}^{\mathtt{u}}$. By defining $\mathbf{h}_\ell^{\mathtt{u}} \triangleq \big[(\mathbf{h}_{1\ell}^{\mathtt{u}})^H,\cdots,(\mathbf{h}_{M\ell}^{\mathtt{u}})^H\big]^H \in \mathbb{C}^{N\times 1}$, $\bar{\mathbf{G}}_{m'}^{\mathtt{AA}} \triangleq \big[(\mathbf{G}_{1m'}^{\mathtt{AA}})^H,\cdots,(\mathbf{G}_{Mm'}^{\mathtt{AA}})^H\big]^H \in \mathbb{C}^{N\times N_{m'}}$, $\mathbf{a}_\ell = [\mathbf{a}_{1\ell},\cdots,\mathbf{a}_{M\ell}] \in \mathbb{C}^{1\times N}$ and $\mathbf{n} \triangleq [\mathbf{n}_1^H,\cdots,\mathbf{n}_M^H]^H \in \mathbb{C}^{N\times 1}$, the SINR in decoding $\mathtt{U}_\ell^{\mathtt{u}}$'s message is given as

$$\gamma_\ell^{\mathtt{u}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha}) = \frac{p_\ell|\mathbf{a}_\ell\mathbf{h}_\ell^{\mathtt{u}}|^2}{\sum_{\ell'\in\mathcal{L}\setminus\{\ell\}}p_{\ell'}|\mathbf{a}_\ell\mathbf{h}_{\ell'}^{\mathtt{u}}|^2 + \mathcal{I}_\ell^{\mathtt{AA}} + \sigma_{\mathtt{AP}}^2\|\mathbf{a}_\ell\|^2}, \quad (4)$$

where $\mathcal{I}_\ell^{\mathtt{AA}} \triangleq \sum_{m'\in\mathcal{M}}\sum_{k\in\mathcal{K}}\alpha_{km'}|\mathbf{a}_\ell\bar{\mathbf{G}}_{m'}^{\mathtt{AA}}\mathbf{w}_{km'}|^2$ is the aggregation of RSI and IAI.

B. Power Consumption Model

We now present a power consumption model that accounts for power consumption for data transmission and baseband processing, as well as circuit operation [40], [41]. As previously discussed, we introduce new binary variables $\mu_m \in \{0,1\}, \forall m \in \mathcal{M}$ to represent operation statuses of $\mathtt{AP}_m$. In particular, $\mathtt{AP}_m$ is selected to be in the active mode if $\mu_m = 1$, and switched to sleep mode otherwise. With $\boldsymbol{\mu} \triangleq \{\mu_m\}_{\forall m\in\mathcal{M}}$, the total power consumption is generally written as

$$P_{\mathtt{T}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}) = P_{\mathtt{D}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}) + P_{\mathtt{C}}(\boldsymbol{\mu}), \quad (5)$$

where $P_{\mathtt{D}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})$ is the power consumption for data transmission and baseband processing, and $P_{\mathtt{C}}(\boldsymbol{\mu})$ is the power consumption for circuit operation; these are detailed as follows:

- The power consumption $P_{\mathtt{D}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})$ can be subcategorized into three main types as:

$$\begin{aligned}P_{\mathtt{D}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}) = &\underbrace{\sum\nolimits_{m\in\mathcal{M}}\frac{\mu_m}{\nu_{\mathtt{AP}_m}}\sum\nolimits_{k\in\mathcal{K}}\|\mathbf{w}_{km}\|^2+\sum\nolimits_{\ell\in\mathcal{L}}\frac{p_\ell}{\nu_\ell^{\mathtt{u}}}}_{\text{radiated power}}\\ &+\underbrace{B\cdot F_{\mathtt{SE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\cdot P^{\mathtt{bh}}}_{\text{load-dependent power}}\\ &+\underbrace{\sum\nolimits_{m\in\mathcal{M}}\sum\nolimits_{k\in\mathcal{K}}\mu_m\alpha_{km}P_{km}^{\mathtt{d}}+L\sum\nolimits_{m\in\mathcal{M}}\mu_m P_m^{\mathtt{u}}}_{\text{baseband power}}.\end{aligned} \tag{6}$$

  The *radiated power* is the power consumption for the transmitted data between APs and UEs, where $\nu_{\mathtt{AP}_m}\in[0,1]$ and $\nu_\ell^{\mathtt{u}}\in[0,1]$ are the power amplifier (PA) efficiencies at $\mathtt{AP}_m$ and $\mathtt{U}_\ell^{\mathtt{u}}$ depending on the design techniques and operating conditions of the PA [31]. The *load-dependent power* is the power consumption spent to transfer the data between APs and CPU in the backhaul which is proportional to the achievable sum rates [41], where B, $F_{\mathtt{SE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})$ and $P^{\mathtt{bh}}$ are the system bandwidth, total SE and average backhaul traffic power of all links, respectively. The *baseband power* is the required power for data processing, waveform design, sync and precoder/receiver computing for $\mathtt{U}_k^{\mathtt{d}}$ (denoted by $P_{km}^{\mathtt{d}}$) and per-UL-user reception (denoted by $P_m^{\mathtt{u}}$). It is obvious that when $\mu_m=0$, $P_{km}^{\mathtt{d}}=P_m^{\mathtt{u}}=0$.
- The power consumption $P_{\mathtt{C}}(\boldsymbol{\mu})$ can be modeled as

$$\begin{aligned}P_{\mathtt{C}}(\boldsymbol{\mu}) = &\sum_{m\in\mathcal{M}}\mu_m P_{\mathtt{AP}_m}^{\mathtt{a}}+\sum_{m\in\mathcal{M}}(1-\mu_m)P_{\mathtt{AP}_m}^{\mathtt{s}}\\ &+\sum_{m\in\mathcal{M}}P_{\mathtt{AP}_m}^{\mathtt{cir}}+\sum_{k\in\mathcal{K}}P_k^{\mathtt{d,cir}}+\sum_{\ell\in\mathcal{L}}P_\ell^{\mathtt{u,cir}},\end{aligned} \tag{7}$$

  where $P_{\mathtt{AP}_m}^{\mathtt{a}}$ and $P_{\mathtt{AP}_m}^{\mathtt{s}}$ are the fixed powers to keep $\mathtt{AP}_m$ in the active and sleep modes, respectively; $P_{\mathtt{AP}_m}^{\mathtt{cir}}$, $P_k^{\mathtt{d,cir}}$ and $P_\ell^{\mathtt{u,cir}}$ are the powers required for circuit operation at $\mathtt{AP}_m$, DL and UL UEs, respectively.

III. Optimization Problem Design

A. Original Problem Formulation

From (2) and (4), the SE in nats/s/Hz is given as

$$F_{\mathtt{SE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\triangleq\sum_{k\in\mathcal{K}}R\big(\gamma_k^{\mathtt{d}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\big)+\sum_{\ell\in\mathcal{L}}R\big(\gamma_\ell^{\mathtt{u}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\big), \tag{8}$$

where $R(x)\triangleq\ln(1+x)$. The EE in nats/Joule is defined as the ratio between the sum throughput (nats/s) and the total power consumption (Watt):

$$F_{\mathtt{EE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})\triangleq\frac{B\cdot F_{\mathtt{SE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})}{P_{\mathtt{T}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})}. \tag{9}$$

To lighten the notations, the system bandwidth B will be omitted in the derivation of the algorithms in this paper, without affecting the optimal solutions. By introducing the constant $\eta\in\{0,1\}$ for the objective-function selection between the SE and EE, the utility function can be written as

$$F(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})=\eta F_{\mathtt{SE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})+(1-\eta)F_{\mathtt{EE}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}).$$

It is worth mentioning that if η = 1 (η = 0, respectively), we arrive at the SE maximization problem (the EE maximization, respectively).

Assuming perfect CSI between APs and UEs, we study a joint design of power control, AP-DL UE association and AP selection, which is formulated as

$$\begin{aligned}\max_{\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}}\quad & F(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}) && \text{(10a)}\\ \text{s.t.}\quad & \mu_m\in\{0,1\},\ \forall m\in\mathcal{M}, && \text{(10b)}\\ & \alpha_{km}\in\{0,1\},\ \forall k\in\mathcal{K},m\in\mathcal{M}, && \text{(10c)}\\ & \|\mathbf{w}_{km}\|^2\le\alpha_{km}P_{\mathtt{AP}_m}^{\max},\ \forall k\in\mathcal{K},m\in\mathcal{M}, && \text{(10d)}\\ & \sum\nolimits_{k\in\mathcal{K}}\alpha_{km}\|\mathbf{w}_{km}\|^2\le\mu_m P_{\mathtt{AP}_m}^{\max},\ \forall m\in\mathcal{M}, && \text{(10e)}\\ & 0\le p_\ell\le P_\ell^{\max},\ \forall\ell\in\mathcal{L}, && \text{(10f)}\\ & R\big(\gamma_k^{\mathtt{d}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\big)\ge\bar{R}_k^{\mathtt{d}},\ \forall k\in\mathcal{K}, && \text{(10g)}\\ & R\big(\gamma_\ell^{\mathtt{u}}(\mathbf{w},\mathbf{p},\boldsymbol{\alpha})\big)\ge\bar{R}_\ell^{\mathtt{u}},\ \forall\ell\in\mathcal{L}. && \text{(10h)}\end{aligned}$$

Constraint (10d) is used to express the AP-UE association, while constraints (10e) and (10f) imply that the transmit powers at $\mathtt{AP}_m$ and $\mathtt{U}_k^{\mathtt{d}}$ are limited by their maximum power budgets $P_{\mathtt{AP}_m}^{\max}$ and $P_\ell^{\max}$, respectively. Moreover, constraints (10g) and (10h) are used to ensure the predetermined rate requirements $\bar{R}_k^{\mathtt{d}}$ and $\bar{R}_k^{\mathtt{u}}$ for $\mathtt{U}_k^{\mathtt{d}}$ and $\mathtt{U}_\ell^{\mathtt{u}}$, respectively. We can see that the objective (10a) is nonconcave and the feasible set is also nonconvex. Hence, problem (10) is a mixed-integer nonconvex optimization problem due to binary variables involved, which is generally NP-hard.

B. Tractable Problem Formulation for (10)

The major difficulty in solving problem (10) is due to binary variables involved. It is not practical to try all possible AP-DL UE associations and AP selections, especially for networks of large size. In addition, the strong coupling between continuous variables $(\mathbf{w},\mathbf{p})$ and binary variables $(\boldsymbol{\alpha},\boldsymbol{\mu})$ makes problem (10) even more difficult. Consequently, the problem is intractable and it is impossible to solve it directly. Even for a fixed $(\boldsymbol{\alpha},\boldsymbol{\mu})$, a direct application of the well-known Dinkelbach algorithm [38] for (10) still involves a nonconvex problem, and thus, its convergence may not be always guaranteed [7]. In what follows, we present a tractable form of (10) by exploiting the special relationship between continuous and binary variables, based on which the combination of ICA method and Dinkelbach transformation can be applied to solve it efficiently for various transmission strategies.

1) Binary Reduction of $\boldsymbol{\alpha}$

The binary variables $\boldsymbol{\alpha}$ and the continuous variables $\mathbf{w}$ are strongly coupled, as revealed by the following lemma.

Lemma 1: If problem (10) contains the optimal solution $\alpha_{km}^*=0$ for some (k,m), it also admits $\mathbf{w}_{km}=\mathbf{0}$ as an

optimal solution for the corresponding beamforming vector.

Proof: Please see Appendix A. ■

From **Lemma** 1, it is straightforward to see that constraint (10d) is naturally satisfied at the optimal point. In particular, when $\alpha_{km}=1$ (or $\alpha_{km}=0$), constraint (10d) is addressed by tighter constraint (10e) (or by **Lemma** 1). In connection to **Lemma** 1, we establish the following main result.

Theorem 1: For any $k\in\mathcal{K}$ and $m\in\mathcal{M}$, let the null space (including zero vector) of $\mathbf{h}_{km}^{\mathtt{d}}$ be $\ker(\mathbf{h}_{km}^{\mathtt{d}})$ and $\mathbf{u}\in\mathcal{U}\triangleq\big\{(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu})|\mathbf{w}_{km}\in\ker(\mathbf{h}_{km}^{\mathtt{d}})\big\}\subseteq\mathcal{F}$, where $\mathcal{F}$ is the feasible set of (10). The following state is obtained:

$$\tilde{\mathbf{u}}=\underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmax}}\ F(\mathbf{u})=(\mathbf{w},\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\mu}|\mathbf{w}_{km}=\mathbf{0}\ \&\ \alpha_{km}=0). \quad (11)$$

Proof: Please see Appendix B. ■

The merits of **Theorem** 1 are detailed as follows. First, even some possible values of $\mathbf{w}_{km}\in\ker\big(\mathbf{h}_{km}^{\mathtt{d}}\big)$ can make $|\mathbf{h}_{km}^{\mathtt{d}}\mathbf{w}_{km}|^2$ to be null, the zero vector $\mathbf{w}_{km}=\mathbf{0}$ is the best value among them. Second, there exists only one of two pairs for the beamforming vector and AP-DL UE association in the optimal solution, which is $(\mathbf{w}_{km}^*,\alpha_{km}^*)\in\big\{(\mathbf{0},0),(\bar{\mathbf{w}},1)\big\}$ with $\bar{\mathbf{w}}\notin\ker\big(\mathbf{h}_{km}^{\mathtt{d}}\big)$. Without loss of optimality, we can replace $\boldsymbol{\alpha}$ by $\mathbf{1}$ in the component containing the compound of $\boldsymbol{\alpha}$ and $\mathbf{w}$, and use a substituting function of $\mathbf{w}$ for $\boldsymbol{\alpha}$ in others. Particularly, we define the 2-tuple of continuous variables as $\mathcal{C}\triangleq\{\mathbf{w},\mathbf{p}\}$, and $\boldsymbol{\Gamma}_{\mathtt{d}}\triangleq\big\{\gamma_k^{\mathtt{d}}(\mathcal{C},\mathbf{1})|\forall k\in\mathcal{K}\big\}$ and $\boldsymbol{\Gamma}_{\mathtt{u}}\triangleq\big\{\gamma_\ell^{\mathtt{u}}(\mathcal{C},\mathbf{1})|\forall\ell\in\mathcal{L}\big\}$ with all entries of $\boldsymbol{\alpha}$ being replaced by ones.

In short, problem (10) can be rewritten as

$$\begin{aligned}
\max_{\mathcal{C}\triangleq\{\mathbf{w},\mathbf{p}\},\boldsymbol{\mu}}\quad & \eta\bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}}\big)+(1-\eta)\bar{F}_{\mathtt{EE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}},\mathcal{C},\boldsymbol{\mu}\big) && \text{(12a)}\\
\text{s.t.}\quad & \sum_{k\in\mathcal{K}}\|\mathbf{w}_{km}\|^2\le\mu_m P_{\mathtt{AP}_m}^{\max},\ \forall m\in\mathcal{M}, && \text{(12b)}\\
& R\big(\gamma_k^{\mathtt{d}}(\mathcal{C},\mathbf{1})\big)\ge\bar{R}_k^{\mathtt{d}},\ \forall k\in\mathcal{K}, && \text{(12c)}\\
& R\big(\gamma_\ell^{\mathtt{u}}(\mathcal{C},\mathbf{1})\big)\ge\bar{R}_\ell^{\mathtt{u}},\ \forall\ell\in\mathcal{L}, && \text{(12d)}\\
& \text{(10b), (10f)}, && \text{(12e)}
\end{aligned}$$

where $\bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}}\big)\triangleq R_\Sigma(\boldsymbol{\Gamma}_{\mathtt{d}})+R_\Sigma(\boldsymbol{\Gamma}_{\mathtt{u}})$, $\bar{F}_{\mathtt{EE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}},\mathcal{C},\boldsymbol{\mu}\big)\triangleq\frac{\bar{F}_{\mathtt{SE}}\left(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}}\right)}{\bar{P}_{\mathtt{T}}(\mathcal{C},\boldsymbol{\mu})}$, and $\bar{P}_{\mathtt{T}}(\mathcal{C},\boldsymbol{\mu})\triangleq P_{\mathtt{T}}(\mathcal{C},f_{\mathtt{spr}}(\mathbf{w}),\boldsymbol{\mu})$ with $R_\Sigma(\mathcal{X})=\sum_{x\in\mathcal{X}}R(x)$. The signal-power ratio function is defined as

$$f_{\mathtt{spr}}:\mathbf{w}\to\mathbf{r}^{\mathtt{sp}}\triangleq\big[r_{\mathtt{sp}}\big(\mathbf{w}_{km},\mathbf{h}_{km}^{\mathtt{d}}|\mathbf{w}_k^{(\kappa)},\mathbf{h}_k^{\mathtt{d}}\big)\big]_{\forall k\in\mathcal{K},m\in\mathcal{M}}, \quad (13)$$

with $\mathbf{h}_k^{\mathtt{d}}\triangleq[\mathbf{h}_{k1}^{\mathtt{d}},\cdots,\mathbf{h}_{kM}^{\mathtt{d}}]\in\mathbb{C}^{1\times N}$, and

$$r_{\mathtt{sp}}\big(\mathbf{x}_1,\mathbf{c}_1|\mathbf{x}_2,\mathbf{c}_2\big)\triangleq\frac{|\mathbf{c}_1\mathbf{x}_1|^2}{|\mathbf{c}_2\mathbf{x}_2|^2+\epsilon}, \quad (14)$$

where ϵ is a very small real number added to avoid a numerical problem when $\mathtt{AP}_m$ turns to sleep mode, and $\mathbf{w}_k^{(\kappa)}$ is a feasible point of $\mathbf{w}_k$ at the κ-th iteration of an iterative algorithm presented shortly. α_{km} is correspondingly replaced by $r_{\mathtt{sp}}\big(\mathbf{w}_{km},\mathbf{h}_{km}^{\mathtt{d}}|\mathbf{w}_k^{(\kappa)},\mathbf{h}_k^{\mathtt{d}}\big)$ in $P_{\mathtt{T}}(\mathcal{C},f_{\mathtt{spr}}(\mathbf{w}),\boldsymbol{\mu})$. We note that (14) is considered as a soft converter from the binary variables into continuous ones, which also indicates the quality of connection between an AP and a UE. As a result, $\mathbf{0}\preceq\mathbf{r}^{\mathtt{sp}}\preceq\mathbf{1}$ is considered as an estimate of $\boldsymbol{\alpha}$ after solving problem (12), i.e.,

$$\alpha_{km}^*=\mathcal{B}\big(r_{\mathtt{sp}}\big(\mathbf{w}_{km}^*,\mathbf{h}_{km}^{\mathtt{d}}|\mathbf{w}_k^*,\mathbf{h}_k^{\mathtt{d}}\big),\varpi\big),\forall k\in\mathcal{K},m\in\mathcal{M}, \quad (15)$$

where

$$\mathcal{B}(x,\varpi)\triangleq\begin{cases}1, & \text{if } x>\varpi,\\ 0, & \text{if } x\le\varpi,\end{cases} \quad (16)$$

and the per-AP power signal ratio ϖ is a small number indicating $x\approx0$ if $x\le\varpi$, and $\mathbf{w}_{km}^*$ is the optimal solution of $\mathbf{w}_{km}$.

Remark *1:* By **Lemma** 1, it is true that $r_{\mathtt{sp}}\big(\mathbf{w}_{km}^*,\mathbf{h}_{km}^{\mathtt{d}}|\mathbf{w}_k^*,\mathbf{h}_k^{\mathtt{d}}\big)\le\varpi$ yields $\mathbf{w}_{km}^*\to\mathbf{0}$. Without loss of optimality, we can omit $\boldsymbol{\alpha}$ in the following derivations. □

2) Binary Reduction of $\boldsymbol{\mu}$

The binary variable $\boldsymbol{\mu}$ is mainly involved in (5). Using **Lemma** 1, we have the following result.

Theorem 2: By treating $\mu_m,\forall m\in\mathcal{M}$ as a constant in each iteration, its solution in the next iteration is iteratively updated as:

$$\begin{aligned}
\mu_m^{(\kappa+1)}=\max\Bigg\{&\max_{k\in\mathcal{K}}\mathcal{B}\Big(r_{\mathtt{sp}}\big(\mathbf{w}_{km}^{(\kappa)},\mathbf{h}_{km}^{\mathtt{d}}|\mathbf{w}_k^{(\kappa)},\mathbf{h}_k^{\mathtt{d}}\big),\varpi\Big),\\
&\max_{\ell\in\mathcal{L}}\mathcal{B}\Big(r_{\mathtt{sp}}\big(\sqrt{p_\ell^{(\kappa)}}\mathbf{h}_{m\ell}^{\mathtt{u}},\mathbf{a}_{m\ell}|\sqrt{p_\ell^{(\kappa)}}\mathbf{h}_\ell^{\mathtt{u}},\mathbf{a}_\ell\big),\varpi\Big)\Bigg\}, \quad (17)
\end{aligned}$$

where the function $\mathcal{B}(:,:)$ was defined in (16).

Proof: Please see Appendix C. ■

From **Theorem 2**, the total power consumption in (5) can be rewritten as

$$\bar{P}_{\mathtt{T}}(\mathcal{C},\boldsymbol{\mu}^{(\kappa)})=P_{\mathtt{D}}(\mathcal{C},f_{\mathtt{spr}}(\mathbf{w}),\boldsymbol{\mu}^{(\kappa)})+P_{\mathtt{C}}(\boldsymbol{\mu}^{(\kappa)}), \quad (18)$$

which involves the continuous variables in $\mathcal{C}$ only.

In summary, the original problem (10) can be cast as the following simplified problem:

$$\begin{aligned}
\max_{\mathcal{C}\triangleq\{\mathbf{w},\mathbf{p}\}}\quad & \eta\bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}}\big)+(1-\eta)\bar{F}_{\mathtt{EE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}},\boldsymbol{\Gamma}_{\mathtt{u}},\mathcal{C},\boldsymbol{\mu}^{(\kappa)}\big) && \text{(19a)}\\
\text{s.t.}\quad & \sum\nolimits_{k\in\mathcal{K}}\|\mathbf{w}_{km}\|^2\le\mu_m^{(\kappa)}P_{\mathtt{AP}_m}^{\max},\ \forall m\in\mathcal{M}, && \text{(19b)}\\
& \text{(10f), (12c), (12d)}. && \text{(19c)}
\end{aligned}$$

Remark *2:* It is clear from the discussion above that solving (19) boils down to finding a saddle point for $\mathcal{C}\triangleq\{\mathbf{w},\mathbf{p}\}$, while the binary variables $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ are post-updated by (15) and (17), respectively. We should emphasize that the binary variables are relaxed to soft-update functions in (19) to reduce the complexity, while maintaining their roles as in the original problem (10). These results hold true for arbitrary linear precoder/receiver schemes, which are discussed in detail next. □

IV. Proposed Solution Based on Zero-Forcing

In this section, we first present an efficient transmission design; this retains the simplicity of the well-known ZF method while enjoys the similar performance of the optimal MMSE method as in the context of massive MIMO [30]. Then, an iterative algorithm based on the ICA method and Dinkelbach transformation is developed to solve the problem design, followed by the initialization discussion.

A. ZF-Based Transmission Design

To make ZF feasible, the total number of APs' antennas is required to be larger than the number of UEs, i.e., $N > \max\{K, L\}$, which can be easily satisfied in massive MIMO systems. For ease of presentation, we first rearrange the sets of beamforming vectors, channel vectors and power allocation between transceivers as follows: $\mathbf{W} \triangleq [\mathbf{w}_1, \cdots, \mathbf{w}_K] \in \mathbb{C}^{N\times K}$, $\mathbf{H}^{\mathtt{d}} \triangleq \left[(\mathbf{h}_1^{\mathtt{d}})^H, \cdots, (\mathbf{h}_M^{\mathtt{d}})^H\right]^H \in \mathbb{C}^{K\times N}$, $\mathbf{H}^{\mathtt{u}} \triangleq \left[\mathbf{h}_1^{\mathtt{u}}, \cdots, \mathbf{h}_L^{\mathtt{u}}\right] \in \mathbb{C}^{N\times L}$, $\mathbf{G}^{\mathtt{cci}} \triangleq \left[(\mathbf{g}_1^{\mathtt{cci}})^H, \cdots, (\mathbf{g}_K^{\mathtt{cci}})^H\right]^H \in \mathbb{C}^{K\times L}$ with $\mathbf{g}_k^{\mathtt{cci}} \triangleq \left[g_{k1}^{\mathtt{cci}}, \cdots, g_{kL}^{\mathtt{cci}}\right] \in \mathbb{C}^{1\times L}$, $\tilde{\mathbf{G}}^{\mathtt{AA}} \triangleq \left[\bar{\mathbf{G}}_1^{\mathtt{AA}}, \cdots, \bar{\mathbf{G}}_M^{\mathtt{AA}}\right] \in \mathbb{C}^{N\times N}$, and $\mathbf{D}^{\mathtt{u}} \triangleq \mathtt{diag}\big(\left[\sqrt{p_1}\cdots\sqrt{p_L}\right]\big)$.

1) ZF-Based DL Transmission

For $\mathbf{H}^{\mathtt{ZF}} = (\mathbf{H}^{\mathtt{d}})^H\big(\mathbf{H}^{\mathtt{d}}(\mathbf{H}^{\mathtt{d}})^H\big)^{-1}$, the ZF precoder matrix is simply computed as $\mathbf{W}^{\mathtt{ZF}} = \mathbf{H}^{\mathtt{ZF}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}}$, where $\mathbf{D}^{\mathtt{d}} = \mathtt{diag}\big(\left[\omega_1\cdots\omega_K\right]\big)$ and ω_k represents the weight for $\mathtt{U}_k^{\mathtt{d}}$. As a result, constraint (19b) becomes

$$\mathtt{Tr}\big((\mathbf{H}^{\mathtt{ZF}})^H\mathbf{B}_m\mathbf{H}^{\mathtt{ZF}}\mathbf{D}^{\mathtt{d}}\big) \le \mu_m^{(\kappa)}P_{\mathtt{AP}_m}^{\max},\ \forall m \in \mathcal{M}, \tag{20}$$

which is a linear constraint, where $\mathbf{B}_m = \mathtt{diag}(\mathbf{b}_m) \in \{0,1\}^{N\times N}$ with

$$\mathbf{b}_m = \big(\underbrace{0\cdots0}_{\sum_{m'=1}^{m-1}N_{m'}}\ \underbrace{1\cdots1}_{N_m}\ 0\cdots0\big). \tag{21}$$

The simplicity of ZF is attributed to the fact that the size of NK scalar variables of $\mathbf{w}$ is now reduced to K scalar variables of $\boldsymbol{\omega} \triangleq [\omega_1, \cdots, \omega_K]^T \in \mathbb{R}^{K\times 1}$. The SINR of $\mathtt{U}_k^{\mathtt{d}}$ with ZF precoder is

$$\gamma_k^{\mathtt{d},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) = \frac{\omega_k|\mathbf{h}_k^{\mathtt{d}}\mathbf{h}_k^{\mathtt{ZF}}|^2}{\|\mathbf{g}_k^{\mathtt{cci}}\mathbf{D}^{\mathtt{u}}\|^2 + \sigma_k^2}, \tag{22}$$

where $\mathbf{h}_k^{\mathtt{ZF}}$ is the k-th column of the ZF precoder $\mathbf{H}^{\mathtt{ZF}}$ and the MUI term $|\mathbf{h}_k^{\mathtt{d}}\mathbf{w}_{k'}|^2 \approx 0, \forall k' \in \mathcal{K}\setminus\{k\}$.

Remark *3:* The following result characterizes the relationship between $\boldsymbol{\omega}$ and $\mathbf{W}^{\mathtt{ZF}}$:

$$\mathbf{W}^{\mathtt{ZF}} = f_{\mathbf{W}}(\boldsymbol{\omega}, \mathbf{H}^{\mathtt{ZF}}), \tag{23}$$

where $f_{\mathbf{W}}(\boldsymbol{\omega}, \mathbf{X}) \triangleq \mathbf{X}\big(\mathtt{diag}(\boldsymbol{\omega})\big)^{\frac{1}{2}}$. Hence, $\mathbf{w}_{km}$ is recovered by extracting from the $((m-1)N_m+1)$-th to (mN_m)-th elements of $\mathbf{w}_k$, where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}^{\mathtt{ZF}}$. □

2) ZF-Based UL Transmission

Let $\mathbf{A}^{\mathtt{ZF}} = \big((\mathbf{H}^{\mathtt{u}})^H\mathbf{H}^{\mathtt{u}}\big)^{-1}(\mathbf{H}^{\mathtt{u}})^H \in \mathbb{C}^{L\times N}$ be the ZF receiver matrix at the CPU. The SINR of $\mathtt{U}_\ell^{\mathtt{u}}$ with ZF receiver is

$$\gamma_\ell^{\mathtt{u},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) = \frac{p_\ell|\mathbf{a}_\ell^{\mathtt{ZF}}\mathbf{h}_\ell^{\mathtt{u}}|^2}{\underbrace{\|\mathbf{a}_\ell^{\mathtt{ZF}}\tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{W}^{\mathtt{ZF}}\|^2}_{\text{IAI + RSI}} + \sigma_{\mathtt{AP}}^2\|\mathbf{a}_\ell^{\mathtt{ZF}}\|^2}, \tag{24}$$

where $\mathbf{a}_\ell^{\mathtt{ZF}}$ is the ℓ-th row of $\mathbf{A}^{\mathtt{ZF}}$.

B. Proposed Algorithm

Before proceeding, we provide some useful approximate functions following ICA properties [36], [37], which will be frequently employed to devise the proposed solutions.

- Consider the convex function $h_{\mathtt{fr}}(x,y) \triangleq x^2/y$ with $(x,y) \in \mathbb{R}_{++}^2$. The concave lower bound of $h_{\mathtt{fr}}(x,y)$ at the feasible point $(x^{(\kappa)}, y^{(\kappa)})$ is given as [19]

$$h_{\mathtt{fr}}(x,y) \ge \frac{2x^{(\kappa)}}{y^{(\kappa)}}x - \frac{(x^{(\kappa)})^2}{(y^{(\kappa)})^2}y := h_{\mathtt{fr}}^{(\kappa)}(x,y). \tag{25}$$

- For the quadratic function $h_{\mathtt{qu}}(z) \triangleq z^2$ with $z \in \mathbb{R}_{++}$, its concave lower bound at $z^{(\kappa)}$ is

$$h_{\mathtt{qu}}(z) \ge 2z^{(\kappa)}z - (z^{(\kappa)})^2 := h_{\mathtt{qu}}^{(\kappa)}(z). \tag{26}$$

Next, problem (19) with ZF design now reduces to the following problem

$$\max_{\boldsymbol{\omega},\mathbf{p}}\ \eta\bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}\big) + (1-\eta)\bar{F}_{\mathtt{EE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}, \mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}\big) \tag{27a}$$

$$\text{s.t.}\quad R\big(\gamma_k^{\mathtt{d},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p})\big) \ge \bar{R}_k^{\mathtt{d}},\ \forall k \in \mathcal{K}, \tag{27b}$$

$$R\big(\gamma_\ell^{\mathtt{u},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p})\big) \ge \bar{R}_\ell^{\mathtt{u}},\ \forall \ell \in \mathcal{L}, \tag{27c}$$

$$(10\text{f}), (20), \tag{27d}$$

where $\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}} \triangleq \{\gamma_k^{\mathtt{d},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) | \forall k \in \mathcal{K}\}$ and $\boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}} \triangleq \{\gamma_\ell^{\mathtt{u},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) | \forall \ell \in \mathcal{L}\}$. Problem (27) involves the nonconcave objective (27a), and nonconvex constraints (27b) and (27c). To apply ICA method, a new transformation with an equivalent feasible set is required. Let us start by rewriting the objective (27a) as $\bar{F}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}, \mathcal{C}^{\mathtt{ZF}}\big) \triangleq \bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}\big)\bar{P}(\mathcal{C}^{\mathtt{ZF}})$, where $\bar{P}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}) \triangleq \big(\eta + \frac{(1-\eta)}{\bar{P}_{\mathtt{T}}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)})}\big)$ and $\mathcal{C}^{\mathtt{ZF}} \triangleq \{\boldsymbol{\omega}, \mathbf{p}\}$.

Theorem 3: Problem (27) is equivalent to the following problem

$$\max_{\boldsymbol{\omega},\mathbf{p},\boldsymbol{\lambda},\phi}\ \frac{\tilde{F}_{\mathtt{SE}}\big(\boldsymbol{\Lambda}_{\mathtt{d}}, \boldsymbol{\Lambda}_{\mathtt{u}}\big)}{\phi} \tag{28a}$$

$$\text{s.t.}\quad \bar{P}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}) \ge 1/\phi, \tag{28b}$$

$$\gamma_k^{\mathtt{d},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) \ge \lambda_k^{\mathtt{d}},\ \forall k \in \mathcal{K}, \tag{28c}$$

$$\gamma_\ell^{\mathtt{u},\mathtt{ZF}}(\boldsymbol{\omega}, \mathbf{p}) \ge \lambda_\ell^{\mathtt{u}},\ \forall \ell \in \mathcal{L}, \tag{28d}$$

$$\lambda_k^{\mathtt{d}} + 1 \ge \exp(\bar{R}_k^{\mathtt{d}}),\ \forall k \in \mathcal{K}, \tag{28e}$$

$$\lambda_\ell^{\mathtt{u}} + 1 \ge \exp(\bar{R}_\ell^{\mathtt{u}}),\ \forall \ell \in \mathcal{L}, \tag{28f}$$

$$(10\text{f}), (20), \tag{28g}$$

where $\tilde{F}_{\mathtt{SE}}\big(\boldsymbol{\Lambda}_{\mathtt{d}}, \boldsymbol{\Lambda}_{\mathtt{u}}\big) \triangleq \ln|\mathbf{I} + \boldsymbol{\Lambda}_{\mathtt{d}}| + \ln|\mathbf{I} + \boldsymbol{\Lambda}_{\mathtt{u}}|$ is a concave function, with $\boldsymbol{\Lambda}_{\mathtt{d}} \triangleq \mathtt{diag}([\lambda_1^{\mathtt{d}}\cdots\lambda_K^{\mathtt{d}}])$ and $\boldsymbol{\Lambda}_{\mathtt{u}} \triangleq \mathtt{diag}([\lambda_1^{\mathtt{u}}\cdots\lambda_L^{\mathtt{u}}])$; ϕ and $\boldsymbol{\lambda} \triangleq \{\boldsymbol{\lambda}_{\mathtt{d}}, \boldsymbol{\lambda}_{\mathtt{u}}\}$ with $\boldsymbol{\lambda}_{\mathtt{d}} \triangleq \{\lambda_k^{\mathtt{d}}\}_{\forall k\in\mathcal{K}}$ and $\boldsymbol{\lambda}_{\mathtt{u}} \triangleq \{\lambda_\ell^{\mathtt{d}}\}_{\forall \ell\in\mathcal{L}}$ are newly introduced variables. Here $\lambda_k^{\mathtt{d}}$ and $\lambda_\ell^{\mathtt{d}}$ can be viewed as soft SINRs for $\mathtt{U}_k^{\mathtt{d}}$ and $\mathtt{U}_\ell^{\mathtt{u}}$, respectively.

Proof: Please see Appendix D. ■

In problem (28), the nonconvex parts include (28b)-(28d). Following the spirit of the ICA method, we introduce a new variable $\xi \in \mathbb{R}_{++}$ to equivalently split constraint (28b) into two constraints as

$$(28\text{b}) \Leftrightarrow \begin{cases} \bar{P}_{\mathtt{T}}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}) \le h_{\mathtt{qu}}(\xi) \triangleq \xi^2, & (29\text{a}) \\ \xi^2 \le \phi(\eta\bar{P}_{\mathtt{T}}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}) + 1 - \eta). & (29\text{b}) \end{cases}$$

We note that $\bar{P}_{\mathtt{T}}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)})$ is a linear function in $(\boldsymbol{\omega}, \mathbf{p})$ due to (23), and thus, (29b) is a second order cone (SOC) representative [42, Sec. 3.3]. Using (26), the nonconvex constraint

(29a) can be iteratively replaced by the following convex one

$$\bar{P}_{\mathtt{T}}(\mathcal{C}^{\mathtt{ZF}}, \boldsymbol{\mu}^{(\kappa)}) \leq h_{\mathtt{qu}}^{(\kappa)}(\xi). \tag{30}$$

Next, we can rewrite (28c) equivalently as

$$(28c) \Leftrightarrow \begin{cases} h_{\mathtt{fr}}(\sqrt{\omega_k}, \psi_k^{\mathtt{d}}) \triangleq \omega_k/\psi_k^{\mathtt{d}} \geq \lambda_k^{\mathtt{d}}, \ \forall k \in \mathcal{K}, & (31a) \\ \psi_k^{\mathtt{d}} \geq \dfrac{\|\mathbf{g}_k^{\mathtt{cci}}\mathbf{D}^{\mathtt{u}}\|^2 + \sigma_k^2}{|\mathbf{h}_k^{\mathtt{d}}\mathbf{h}_k^{\mathtt{ZF}}|^2}, \forall k \in \mathcal{K}, & (31b) \end{cases}$$

where $\boldsymbol{\psi}^{\mathtt{d}} \triangleq \{\psi_k^{\mathtt{d}}\}_{\forall k \in \mathcal{K}}$ are new variables. Since (31b) is a linear constraint, we focus on convexifying (31a) using (25) as

$$h_{\mathtt{fr}}^{(\kappa)}(\sqrt{\omega_k}, \psi_k^{\mathtt{d}}) \geq \lambda_k^{\mathtt{d}}, \ \forall k \in \mathcal{K}. \tag{32}$$

Similarly, with new variables $\boldsymbol{\psi}^{\mathtt{u}} \triangleq \{\psi_\ell^{\mathtt{u}}\}_{\forall \ell \in \mathcal{L}}$, constraint (28d) is iteratively replaced by the following two convex ones

$$h_{\mathtt{fr}}^{(\kappa)}(\sqrt{p_\ell}, \psi_\ell^{\mathtt{u}}) \geq \lambda_\ell^{\mathtt{u}}, \ \forall \ell \in \mathcal{L}, \tag{33a}$$

$$\psi_\ell^{\mathtt{u}} \geq \frac{\|\mathbf{a}_\ell^{\mathtt{ZF}}\tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{W}^{\mathtt{ZF}}\|^2 + \sigma_{\mathtt{AP}}^2\|\mathbf{a}_\ell^{\mathtt{ZF}}\|^2}{|\mathbf{a}_\ell^{\mathtt{ZF}}\mathbf{h}_\ell^{\mathtt{u}}|^2}, \forall \ell \in \mathcal{L}. \tag{33b}$$

With the above discussions based on the ICA method, we obtain the following approximate problem of (28) (and hence (27)) with the convex set solved at iteration $(\kappa+1)$:

$$\max_{\boldsymbol{\omega}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \xi, \phi} \quad \frac{\tilde{F}_{\mathtt{SE}}(\boldsymbol{\Lambda}_{\mathtt{d}}, \boldsymbol{\Lambda}_{\mathtt{u}})}{\phi} \tag{34a}$$

$$\text{s.t.} \quad (10f), (20), (28e), (28f), (29b), (30), (31b), (32), (33), \tag{34b}$$

where $\boldsymbol{\psi} \triangleq \{\boldsymbol{\psi}^{\mathtt{d}}, \boldsymbol{\psi}^{\mathtt{u}}\}$. We can see that the set of variables in (34) is independent of the numbers of APs and antennas, which differs from the original problem (10). The objective (34a) is a concave-over-linear function, which can be efficiently addressed by the Dinkelbach transformation. In particular, we have

$$\mathcal{V}^{\mathtt{ZF},(\kappa+1)} = \underset{\mathcal{V}^{\mathtt{ZF}} \in \mathcal{F}^{(\kappa)}}{\operatorname{argmax}} \ \ddot{F}^{(\kappa)} \triangleq \tilde{F}_{\mathtt{SE}}(\boldsymbol{\Lambda}_{\mathtt{d}}, \boldsymbol{\Lambda}_{\mathtt{u}}) - t^{(\kappa)}\phi, \tag{35}$$

where $\mathcal{F}^{(\kappa)} \triangleq \big\{\mathcal{V}^{\mathtt{ZF}} \triangleq \{\boldsymbol{\omega}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \xi, \phi\} | (34b) \text{ holds}\big\}$ and $t^{(\kappa)} \triangleq \frac{\tilde{F}_{\mathtt{SE}}(\boldsymbol{\Lambda}_{\mathtt{d}}^{(\kappa)}, \boldsymbol{\Lambda}_{\mathtt{u}}^{(\kappa)})}{\phi^{(\kappa)}}$. To start the computational procedure, an initial feasible point $(\boldsymbol{\omega}^{(0)}, \mathbf{p}^{(0)}, \boldsymbol{\psi}^{(0)}, \xi^{(0)}, \phi^{(0)})$ for (35) is required. This is done by guaranteeing QoS constraints (28e) and (28f) to be satisfied. Thus, we successively solve the following simplified problem of (34)

$$\max_{\mathcal{V}^{\mathtt{ZF}}, \boldsymbol{\theta}} \quad \Theta \triangleq \sum\nolimits_{k \in \mathcal{K}} \theta_k^{\mathtt{d}} + \sum\nolimits_{\ell \in \mathcal{L}} \theta_\ell^{\mathtt{u}} \tag{36a}$$

$$\text{s.t.} \quad \lambda_k^{\mathtt{d}} + 1 - \exp(\bar{R}_k^{\mathtt{d}}) \geq \theta_k^{\mathtt{d}}, \ \forall k \in \mathcal{K}, \tag{36b}$$

$$\lambda_\ell^{\mathtt{u}} + 1 - \exp(\bar{R}_\ell^{\mathtt{u}}) \geq \theta_\ell^{\mathtt{u}}, \ \forall \ell \in \mathcal{L}, \tag{36c}$$

$$\theta_k^{\mathtt{d}} \leq 0, \ \forall k \in \mathcal{K}, \ \ \theta_\ell^{\mathtt{u}} \leq 0, \forall \ell \in \mathcal{L}, \tag{36d}$$

$$(10f), (20), (29b), (30), (31b), (32), (33), \tag{36e}$$

where $\boldsymbol{\theta} \triangleq \{\theta_k^{\mathtt{d}}, \theta_\ell^{\mathtt{u}}\}_{\forall k \in \mathcal{K}, \ell \in \mathcal{L}}$ are slack variables. The initial feasible point $(\boldsymbol{\omega}^{(0)}, \mathbf{p}^{(0)}, \boldsymbol{\psi}^{(0)}, \xi^{(0)})$ for (35) is found when the objective (36a) is close to zero, and $\phi^{(0)} = 1/\bar{P}(\mathcal{C}^{\mathtt{ZF},(0)}, \boldsymbol{\mu}^{(0)})$. The proposed algorithm for solving the ZF-based SE-EE problem (10) is summarized in Algorithm 1.

Algorithm 1 Proposed Algorithm to Solve ZF-based SE-EE Problem (10)

1: **Initialization:** Compute ZF precoder and receiver: $\mathbf{H}^{\mathtt{ZF}}$ and $\mathbf{A}^{\mathtt{ZF}}$.
2: Set $\ddot{F}^{(\kappa)} := -\infty$, $\kappa := 0$, and solve (36) to generate an initial feasible point $(\boldsymbol{\omega}^{(0)}, \mathbf{p}^{(0)}, \boldsymbol{\psi}^{(0)}, \xi^{(0)}, \phi^{(0)})$.
3: **repeat** {Solving (27)}
4: Solve (35) to obtain the optimal solution $(\boldsymbol{\omega}^\star, \mathbf{p}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\psi}^\star, \xi^\star, \phi^\star)$ and $\ddot{F}^{(\kappa+1)}$.
5: Update $(\boldsymbol{\omega}^{(\kappa+1)}, \mathbf{p}^{(\kappa+1)}, \boldsymbol{\psi}^{(\kappa+1)}, \xi^{(\kappa+1)}, \phi^{(\kappa+1)}) := (\boldsymbol{\omega}^\star, \mathbf{p}^\star, \boldsymbol{\psi}^\star, \xi^\star, \phi^\star)$.
6: Set $\kappa := \kappa + 1$.
7: **until** Convergence
8: Update $(\boldsymbol{\omega}^*, \mathbf{p}^*) := (\boldsymbol{\omega}^{(\kappa)}, \mathbf{p}^{(\kappa)})$.
9: Use (23) to recover $\mathbf{w}_{km}, \ \forall k \in \mathcal{K}, m \in \mathcal{M}$.
10: Compute $\boldsymbol{\alpha}^*$ and $\boldsymbol{\mu}^*$ as in (15) and (17), respectively.
11: Repeat Steps 1-9 with fixed values of $\boldsymbol{\alpha}^*$ and $\boldsymbol{\mu}^*$ to find the exact solution of $(\mathbf{w}^*, \mathbf{p}^*)$.
12: Use $(\mathbf{w}^*, \mathbf{p}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*)$ to compute $F(\mathbf{w}^*, \mathbf{p}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*)$ in (10a).
13: **Output:** The optimal solution $(\mathbf{w}^*, \mathbf{p}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*)$ and objective value $F(\mathbf{w}^*, \mathbf{p}^*, \boldsymbol{\alpha}^*, \boldsymbol{\mu}^*)$.

C. Convergence and Complexity Analysis

1) Convergence Analysis

Algorithm 1 is mainly based on inner approximation and Dinkelbach transformation, where their convergences were proved in [36] and [38], respectively. Specifically, as provided in [38], the optimal solution of problem (35) is derived as a minorant obtained at each iteration of the ICA-based approximate problem (34). From the properties of the ICA method [37], it follows that $\mathcal{F}^{(\kappa)} \subset \mathcal{F}^{(\kappa+1)}$, resulting in a sequence $(\boldsymbol{\omega}^{(\kappa)}, \mathbf{p}^{(\kappa)}, \boldsymbol{\psi}^{(\kappa)}, \xi^{(\kappa)}, \phi^{(\kappa)})$ of improved points of (34) and a sequence of non-decreasing objective values. Moreover, $\mathcal{F}^{(\kappa)}$ is a convex connected set, as shown in [17]. Therefore, Algorithm 1 is guaranteed to arrive at least at a locally optimal solution for (28) (and hence (27)) when $\kappa \to \infty$, satisfying the Karush-Kuhn-Tucker conditions according to [36, Theorem 1].

2) Computational Complexity

Before deriving the complexity, we consider the following stages of Algorithm 1:

- The pre-processing stage computes constant matrices, i.e., $\mathbf{H}^{\mathtt{ZF}}$ and $\mathbf{A}^{\mathtt{ZF}}$. This stage contributes a minor part to the total complexity since it only executes the matrix computation, which can be done easily. For the ZF design, it implies a computational complexity of $\mathcal{O}(N^3)$ floating operations (flops).
- The major complexity comes from the optimization of the involved variables. This is associated to the main loop in solving (27) (i.e., Steps 3-7 in Algorithm 1), of which the per-iteration complexity is $\mathcal{O}\big(c^{2.5}(v^2+c)\big)$, with $v = (3K + 3L + 2)$ scalar decision variables and $c = (M + 3K + 4L + 2)$ linear/SOC constraints [17].

It can be observed that the per-iteration complexity for the main loop is less dependent on M, since problem (35) only contains M linear constraints in (20). Moreover, the size of $\mathbf{H}^{\mathtt{ZF}}$ and $\mathbf{A}^{\mathtt{ZF}}$ remains unchanged. Therefore, the complexity based on the proposed design is almost the same for different transmission strategies. In Table I, we provide the major com-

TABLE I
COMPLEXITY COMPARISON.

Transmission strategies	Pre-processing (flops)	Per-iteration complexity for optimization
Proposed ZF-based design	$\mathcal{O}(N^3)$	$\mathcal{O}\big(c^{2.5}(v^2+c)\big)$
MRT/MRC	$\mathcal{O}(N^2)$	$\mathcal{O}\big(c^{2.5}(v^2+c)\big)$

plexities of the proposed ZF and MRT/MRC, which are quite comparative. However, the execution time partially depends on the complexity of solving the successive approximate program in an iterative algorithm, as well as the feasible region under the structure of constant matrices in the pre-processing stage. This will be further elaborated through numerical examples.

V. PROPOSED SOLUTION BASED ON IMPROVED ZERO-FORCING

From (24), it can be seen that the IAI and RSI are still the main limitations of FD CF-mMIMO. Thus, our next endeavor is to propose an IZF-based design to manage the network interference more effectively. In particular, ONB-ZF with PCA, referred to as ONB-ZF-PCA, is developed for the DL transmission to mitigate the effects of IAI and RSI. In addition, we also adopt a ZF-SIC receiver for UL reception to further accelerate the performance of the ZF-based design.

A. IZF-Based Transmission Design

1) ONB-ZF-PCA-Based DL Transmission

The key idea of the ONB-ZF-PCA method is to utilize ONB-ZF for MUI cancellation and exploit PCA to depress the effects of IAI and RSI on UL transmission. For $\mathbf{W}^{\mathtt{IZF}} \triangleq [\mathbf{w}_1 \ \ldots \ \mathbf{w}_K] \in \mathbb{C}^{N\times K}$, we introduce the ONB-ZF-PCA procedure and its operation as follows.

Procedure 1: The ONB-ZF-PCA precoder is computed as

$$\mathbf{W}^{\mathtt{IZF}} = \mathbf{P}\mathbf{Q}^H\tilde{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}}, \tag{37}$$

where $\mathbf{D}^{\mathtt{d}}$ was already defined in (20), and other matrix components $\mathbf{P}$, $\mathbf{Q}$, and $\tilde{\mathbf{T}}$ are determined by the following steps:

1) Using the PCA method, we can express the covariance matrix of $\tilde{\mathbf{G}}^{\mathtt{AA}}$ as
$$(\tilde{\mathbf{G}}^{\mathtt{AA}})^H\tilde{\mathbf{G}}^{\mathtt{AA}} = \mathbf{U}\mathbf{E}\mathbf{U}^H, \tag{38}$$
where $\mathbf{U}$ and $\mathbf{E}$ are unitary and diagonal matrices, respectively, which are derived by using singular value decomposition (SVD).
2) Let $\mathbf{e} \triangleq [e_1, \cdots, e_N]$, where $e_1 \geq e_2 \geq \cdots \geq e_N \geq 0$ are eigenvalues on the diagonal of $\mathbf{E}$. We define $\bar{N}$ as
$$\bar{N} \triangleq \min\Big\{\big\{n \in \mathcal{N} \triangleq \{1,\cdots,N\} \big| f_n^{\mathtt{ER}}(\mathbf{e}) \geq \delta \wedge n < N-1\big\} \cup \{N-1\}\Big\}, \tag{39}$$
where $f_n^{\mathtt{ER}}(\mathbf{e}) = \frac{\sum_{i=1}^{n} e_i}{\sum_{i=1}^{N} e_i}$ denotes the ratio of the first n eigenvalues to the sum of all eigenvalues, and δ is a percentage threshold for the sum of first $\bar{N}$ eigenvalues over all eigenvalues with $\delta = 99\%$.
3) **Compute $\mathbf{P}$**: We compute $\mathbf{P} = \mathbf{I} - \bar{\mathbf{U}}\bar{\mathbf{U}}^H$, where $\bar{\mathbf{U}}$ is generated from the first $\bar{N}$ columns of $\mathbf{U}$.
4) **Compute $\mathbf{Q}$**: The economy-size LQ decomposition is applied to the compound matrix $\mathbf{H}^{\mathtt{d}}\mathbf{P}$ such as $\mathbf{H}^{\mathtt{d}}\mathbf{P} = \mathbf{T}\mathbf{Q}$, where $\mathbf{T} \in \mathbb{C}^{K\times K}$ is a lower triangular matrix and $\mathbf{Q} \in \mathbb{C}^{K\times N}$ is an ONB matrix. Since $K \ll N$ in CF-mMIMO, the economy-size decomposition can be exploited to reduce the computational complexity, leading to $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$ but $\mathbf{Q}^H\mathbf{Q} \neq \mathbf{I}$.
5) **Compute $\tilde{\mathbf{T}}$**: The entry at the i-th row and j-th column of $\tilde{\mathbf{T}} \in \mathbb{C}^{K\times K}$, denoted by $\tilde{T}_{ij}$, is generally computed by using the following recursive expression:
$$\tilde{T}_{ij} = \begin{cases} -\frac{1}{T_{ii}}\sum_{j'=j}^{i-1} T_{ij'}\tilde{T}_{j'j}, & \text{if } i > j, \\ 1, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases} \tag{40}$$
where T_{ij} denotes the entry at the i-th row and j-th column of $\mathbf{T}$ obtained in Step 4.

Proof: Please see Appendix E. ∎

Remark *4:* We note that the matrix $\mathbf{P}$ computed via the PCA method aims at mitigating the effects of IAI and RSI. On the other hand, we can use the ZF precoder matrix based on ONB only (i.e., by skipping Steps 1-3). The LQ decomposition in Step 4 is applied to $\mathbf{H}^{\mathtt{d}}$ instead of $\mathbf{H}^{\mathtt{d}}\mathbf{P}$, i.e., $\mathbf{H}^{\mathtt{d}} = \mathbf{T}\mathbf{Q}$. Then, a simpler precoder matrix can be constructed as

$$\mathbf{W}^{\mathtt{ONB\text{-}ZF}} = \mathbf{Q}^H\tilde{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}}. \tag{41}$$

Based on **Procedure 1** and in the same manner as (20), the power constraint at $\mathtt{AP}_m$ becomes

$$\mathtt{Tr}\big(\tilde{\mathbf{T}}^H\mathbf{Q}\mathbf{P}\mathbf{B}_m\mathbf{P}^H\mathbf{Q}^H\tilde{\mathbf{T}}\mathbf{D}^{\mathtt{d}}\big) \leq \mu_m^{(\kappa)}P_{\mathtt{AP}_m}^{\max}, \ \forall m \in \mathcal{M}. \tag{42}$$

The SINR at $\mathtt{U}_k^{\mathtt{d}}$ with the ONB-ZF-PCA precoder is

$$\gamma_k^{\mathtt{d,IZF}}(\boldsymbol{\omega},\mathbf{p}) = \frac{\omega_k|\mathbf{h}_k^{\mathtt{d}}\mathbf{h}_k^{\mathtt{IZF}}|^2}{\|\mathbf{g}_k^{\mathtt{cci}}\mathbf{D}^{\mathtt{u}}\|^2 + \sigma_k^2}, \tag{43}$$

where $\mathbf{h}_k^{\mathtt{IZF}}$ is the k-th column of $\mathbf{H}^{\mathtt{IZF}} \triangleq \mathbf{P}\mathbf{Q}^H\tilde{\mathbf{T}}$.

2) ZF-SIC-Based UL Transmission

The decoded signals are successively removed before decoding the next signals, following the SIC principle [43]. Assuming that the decoding order follows the UL UEs' index, i.e., $\ell = 1, 2, \cdots, L$, the $\mathtt{U}_\ell^{\mathtt{u}}$'s signal is decoded by treating signals of $\mathtt{U}_{\ell'}^{\mathtt{u}}$ for $\ell' \geq \ell+1$ as noise, while other signals are removed by SIC. The remaining MUI at $\mathtt{U}_\ell^{\mathtt{u}}$ is further canceled by the ZF receiver. Thus, the ZF-SIC receiver for decoding $\mathtt{U}_\ell^{\mathtt{u}}$'s signal can be expressed as $\mathbf{a}_\ell^{\mathtt{IZF}}$, which is the first row of $\mathbf{A}_\ell^{\mathtt{IZF}} \triangleq \big((\bar{\mathbf{H}}_\ell^{\mathtt{u}})^H\bar{\mathbf{H}}_\ell^{\mathtt{u}}\big)^{-1}(\bar{\mathbf{H}}_\ell^{\mathtt{u}})^H \in \mathbb{C}^{(L-\ell+1)\times N}$, where $\bar{\mathbf{H}}_\ell^{\mathtt{u}} \triangleq \big[\mathbf{h}_\ell^{\mathtt{u}}, \cdots, \mathbf{h}_L^{\mathtt{u}}\big] \in \mathbb{C}^{N\times(L-\ell+1)}$ and $\bar{\mathbf{D}}^{\mathtt{u}} \triangleq \mathtt{diag}\big(\big[\sqrt{p_\ell}, \cdots, \sqrt{p_L}\big]\big)$. Accordingly, the SINR of $\mathtt{U}_\ell^{\mathtt{u}}$ with ZF-SIC receiver becomes

$$\gamma_\ell^{\mathtt{u,IZF}}(\boldsymbol{\omega},\mathbf{p}) = \frac{p_\ell|\mathbf{a}_\ell^{\mathtt{IZF}}\mathbf{h}_\ell^{\mathtt{u}}|^2}{\|\mathbf{a}_\ell^{\mathtt{IZF}}\tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{W}^{\mathtt{IZF}}\|^2 + \sigma_{\mathtt{AP}}^2\|\mathbf{a}_\ell^{\mathtt{IZF}}\|^2}, \tag{44}$$

where $\|\mathbf{a}_\ell^{\mathtt{IZF}}\bar{\mathbf{H}}_{\ell+1}^{\mathtt{u}}\bar{\mathbf{D}}_{\ell+1}^{\mathtt{u}}\|^2 \to 0$ due to the ZF-SIC matrix $\mathbf{A}_\ell^{\mathtt{IZF}}$.

B. IZF-Based Optimization Problem

Similarly to problem (27), the IZF-based optimization problem can be expressed as

$$\max_{\boldsymbol{\omega},\mathbf{p}} \quad \eta\bar{F}_{\mathtt{SE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{IZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{IZF}}\big) + (1-\eta)\bar{F}_{\mathtt{EE}}\big(\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{IZF}}, \boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{IZF}}, \mathcal{C}^{\mathtt{IZF}}, \boldsymbol{\mu}^{(\kappa)}\big) \tag{45a}$$

$$\text{s.t.} \quad R\big(\gamma_k^{\mathtt{d},\mathtt{IZF}}(\boldsymbol{\omega},\mathbf{p})\big) \geq \bar{R}_k^{\mathtt{d}},\ \forall k \in \mathcal{K}, \tag{45b}$$

$$R\big(\gamma_\ell^{\mathtt{u},\mathtt{IZF}}(\boldsymbol{\omega},\mathbf{p})\big) \geq \bar{R}_\ell^{\mathtt{u}},\ \forall \ell \in \mathcal{L}, \tag{45c}$$

$$(10\text{f}), (42), \tag{45d}$$

where $\boldsymbol{\Gamma}_{\mathtt{d}}^{\mathtt{IZF}} \triangleq \{\gamma_k^{\mathtt{d},\mathtt{IZF}}(\boldsymbol{\omega},\mathbf{p}) | \forall k \in \mathcal{K}\}$, $\boldsymbol{\Gamma}_{\mathtt{u}}^{\mathtt{IZF}} \triangleq \{\gamma_\ell^{\mathtt{u},\mathtt{IZF}}(\boldsymbol{\omega},\mathbf{p}) | \forall \ell \in \mathcal{L}\}$ and $\mathcal{C}^{\mathtt{IZF}} \triangleq \{\boldsymbol{\omega},\mathbf{p}\}$. It can be seen that the nonconvex parts in problems (27) and (45) have a similar structure, and thus, we can slightly modify Algorithm 1 to solve (45).

VI. Proposed Heap-Based Pilot Assignment Strategy

The developments presented in the previous sections are based on the assumption of perfect CSI to realize the potential performance of the proposed FD CF-mMIMO. However, it is of practical interest to take imperfect CSI into account. Each coherence interval in FD CF-mMIMO can be divided into two phases: UL training and data transmission in DL-UL. The coherence interval is short, and thus, each UE should practically be assigned a non-orthogonal pilot sequence, resulting in the well-known pilot contamination problem [29]. Therefore, the main goal of this section is to develop a pilot assignment algorithm based on the heap structure to reduce the effect of pilot contamination and training complexity. We note that the pilot assignment based on greedy method given in [23] not only requires high complexity due to the strategy of trial and error, but also is inapplicable to FD CF-mMIMO due to the additional channel estimation of CCI links.

Remark *5:* The channels of fading loop and IAI (i.e., $\mathbf{G}_{mm}^{\mathtt{SI}}$ and $\mathbf{G}_{mm'}^{\mathtt{AA}}, \forall m \neq m'$) are assumed to be the same as before. The reason for the fading loop channel is that the transmit and receive antennas are co-located at the APs. On the other hand, APs are generally fixed in a given area without mobility, and thus, the IAI channels can be perfectly acquired at the CPU at the initial deployment of the FD CF-mMIMO networks.

A. Channel Estimation and MSE Minimization Problem

We assume that all UEs share the same orthogonal set of pilots, and the DL and UL UEs send training sequences in different intervals to allow the channel estimation of CCI links. Let $\tau < \min\{K, L\}$ be the length of pilot sequences. Then, the pilot set is defined as $\boldsymbol{\Xi} \triangleq [\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_\tau] \in \mathbb{C}^{\tau\times\tau}$, where $\boldsymbol{\varphi}_i \in \mathbb{C}^{\tau\times 1}$ satisfies the orthogonality, i.e., $\boldsymbol{\varphi}_i^H \boldsymbol{\varphi}_{i'} = 1$ if $i = i' \in \mathcal{T}_{\mathtt{p}} \triangleq \{1, \cdots, \tau\}$, and $\boldsymbol{\varphi}_i^H \boldsymbol{\varphi}_{i'} = 0$, otherwise. We introduce the assignment variable $\upsilon_{ij} \in \{0, 1\}$ to determine whether the i-th pilot sequence is assigned to the j-th UE, with $j \in \mathcal{T}_{\mathtt{u}} \triangleq \{1, \cdots, U\}$ and $U \in \{K, L\}$. As a result, the pilot assigned to UE j can be expressed as $\bar{\boldsymbol{\varphi}}_j = \boldsymbol{\varphi}_i$ if $\upsilon_{ij} = 1$. Let $\bar{\boldsymbol{\Xi}} \triangleq [\bar{\boldsymbol{\varphi}}_1, \cdots, \bar{\boldsymbol{\varphi}}_U] \in \mathbb{C}^{\tau\times U}$ be the pilot assignment matrix, such as $\bar{\boldsymbol{\Xi}} = \boldsymbol{\Xi}\boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon} \triangleq [\upsilon_{ij}]_{i\in\mathcal{T}_{\mathtt{p}}, j\in\mathcal{T}_{\mathtt{u}}} \in \mathbb{C}^{\tau\times U}$ following by the condition: $\sum_{i\in\mathcal{T}_{\mathtt{p}}} \upsilon_{ij} \leq 1,\ \forall j \in \mathcal{T}_{\mathtt{u}}$.

The training procedure for FD CF-mMIMO in TDD operation is executed in two phases. In the first phase, UL UEs send their pilot signals to APs to perform the channel estimation, and at the same time DL UEs also receive UL pilots to estimate CCI channels. In the second phase, DL UEs send their pilot signals along with the estimates of CCI links to APs. The training signals received at $\mathtt{AP}_m$ can be written as $\mathbf{Y}_m^{\mathtt{tr}} = \sum_{j'\in\mathcal{T}_{\mathtt{u}}} \sqrt{\tau p_{j'}^{\mathtt{tr}}} \bar{\boldsymbol{\varphi}}_{j'} \mathbf{h}_{mj'} + \mathbf{Z}_m$, where $\mathbf{h}_{mj} \in \{\mathbf{h}_{km}^{\mathtt{d}}, (\mathbf{h}_{m\ell}^{\mathtt{u}})^H\} \in \mathbb{C}^{1\times N_m}$, and $p_j^{\mathtt{tr}}$ and $\mathbf{Z}_m \sim \mathcal{CN}(0, \sigma_{\mathtt{AP}}^2 \mathbf{I})$ denote the UL training power of UE j and the AWGN, respectively. Using the linear MMSE (LMMSE) estimation [44], the channel estimate of $\mathbf{h}_{mj}$ is given as

$$\hat{\mathbf{h}}_{mj} = \frac{\sqrt{\tau p_j^{\mathtt{tr}}}\beta_{mj}}{\sum_{j'\in\mathcal{T}_{\mathtt{u}}} \tau p_{j'}^{\mathtt{tr}} \beta_{mj'} |\bar{\boldsymbol{\varphi}}_j^H \bar{\boldsymbol{\varphi}}_{j'}|^2 + \sigma_{\mathtt{AP}}^2} \bar{\boldsymbol{\varphi}}_j^H \mathbf{Y}_m^{\mathtt{tr}}, \tag{46}$$

where $\beta_{mj} \in \{\beta_{km}^{\mathtt{d}}, \beta_{m\ell}^{\mathtt{u}}\}$ is the large-scale fading of the link between $\mathtt{AP}_m$ and UE j. We denote $\tilde{\mathbf{h}}_{mj} = \mathbf{h}_{mj} - \hat{\mathbf{h}}_{mj}$ as the channel estimation error, which is independent of $\mathbf{h}_{mj}$. The elements of $\tilde{\mathbf{h}}_{mj}$ can be modeled as i.i.d. $\mathcal{CN}(0, \varepsilon_{mj})$ RVs, where

$$\varepsilon_{mj} = \beta_{mj}\Big(1 - \frac{\tau p_j^{\mathtt{tr}} \beta_{mj}}{\sum_{j'\in\mathcal{T}_{\mathtt{u}}} \tau p_{j'}^{\mathtt{tr}} \beta_{mj'} |\bar{\boldsymbol{\varphi}}_j^H \bar{\boldsymbol{\varphi}}_{j'}|^2 + \sigma_{\mathtt{AP}}^2}\Big). \tag{47}$$

In an analogous fashion, the channel estimate and channel estimation error of CCI link $g_{k\ell}^{\mathtt{cci}}$ executed at $\mathtt{U}_k^{\mathtt{d}}$ are given as

$$\hat{g}_{k\ell}^{\mathtt{cci}} = \frac{\sqrt{\tau p_\ell^{\mathtt{tr}}}\beta_{k\ell}^{\mathtt{cci}}}{\sum_{\ell'\in\mathcal{L}} \tau p_{\ell'}^{\mathtt{tr}} \beta_{k\ell'}^{\mathtt{cci}} |\bar{\boldsymbol{\varphi}}_\ell^H \bar{\boldsymbol{\varphi}}_{\ell'}|^2 + \sigma_k^2} \bar{\boldsymbol{\varphi}}_\ell^H \mathbf{y}_k^{\mathtt{tr},\mathtt{cci}}, \tag{48}$$

and $\tilde{g}_{k\ell}^{\mathtt{cci}} \sim \mathcal{CN}(0, \varepsilon_{k\ell}^{\mathtt{cci}})$, respectively, where

$$\varepsilon_{k\ell}^{\mathtt{cci}} = \beta_{k\ell}^{\mathtt{cci}}\Big(1 - \frac{\tau p_\ell^{\mathtt{tr}} \beta_{k\ell}^{\mathtt{cci}}}{\sum_{\ell'\in\mathcal{L}} \tau p_{\ell'}^{\mathtt{tr}} \beta_{k\ell'}^{\mathtt{cci}} |\bar{\boldsymbol{\varphi}}_\ell^H \bar{\boldsymbol{\varphi}}_{\ell'}|^2 + \sigma_k^2}\Big). \tag{49}$$

Here, $\beta_{k\ell}^{\mathtt{cci}}$ denotes the large-scale fading of CCI link $\mathtt{U}_\ell^{\mathtt{u}} \to \mathtt{U}_k^{\mathtt{d}}$, and $\mathbf{y}_k^{\mathtt{tr},\mathtt{cci}} = \sum_{\ell\in\mathcal{T}_{\mathtt{u}}} \sqrt{\tau p_{j'}^{\mathtt{tr}}} \bar{\boldsymbol{\varphi}}_\ell g_{k\ell}^{\mathtt{cci}} + \mathbf{z}_k$, with $\mathbf{z}_k \sim \mathcal{CN}(0, \sigma_k^2 \mathbf{I})$, is the UL UEs' training signals received at $\mathtt{U}_k^{\mathtt{d}}$.

The CSI of CCI links is directly fed back to APs using a dedicated control channel to ensure a low-complexity channel estimation at DL UEs. To mitigate the effects of pilot contamination, a pilot assignment for the main DL and UL channels is far more important that of CCI channels. Thus, we consider the following MSE minimization problem:

$$\min_{\boldsymbol{\Upsilon}} \quad \max_{j\in\mathcal{T}_{\mathtt{u}}} \sum\nolimits_{m\in\mathcal{M}} \frac{N_m \varepsilon_{mj}}{\beta_{mj}} \tag{50a}$$

$$\text{s.t.} \quad \upsilon_{ij} \in \{0, 1\},\ \sum\nolimits_{i\in\mathcal{T}_{\mathtt{p}}} \upsilon_{ij} \leq 1,\ \forall i \in \mathcal{T}_{\mathtt{p}}, \forall j \in \mathcal{T}_{\mathtt{u}}. \tag{50b}$$

B. Heap Structure-Based Pilot Assignment Strategy

Problem (50) is a min-max problem for sum of fractional functions, for which it is hard to find an optimal solution. For an efficient solution, we first introduce the following theorem.

Theorem 4: Problem (50) can be solved via the following tractable problem:

$$\min_{\boldsymbol{\Upsilon}} \quad \max_{j\in\mathcal{T}_{\mathtt{u}}} \sum\nolimits_{j'\in\mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} \boldsymbol{\upsilon}_j^H \boldsymbol{\upsilon}_{j'},\ \text{s.t.} \quad (50\text{b}), \tag{51}$$

where $\tilde{\beta}_{j'} \triangleq \sum_{m\in\mathcal{M}} N_m \tau p_{j'}^{\mathtt{tr}} \beta_{mj'}$.

Proof: Please see Appendix F. ∎

Remark *6:* For $\bar{\upsilon}_{jj'} = \boldsymbol{\upsilon}_j^H \boldsymbol{\upsilon}_{j'}$, the objective function in (51) becomes a kind of bottleneck assignment problem with the partial of max function replaced by $\min \max_{j\in\mathcal{T}_{\mathtt{u}}} \sum_{j'\in\mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} \bar{\upsilon}_{jj'}$. This indicates that all pilots play

the same roles and the optimal solution for pilot assignment depends on how to cluster UEs which share the same training sequence. □

Thus far, we have provided the tractable MSE minimization problem. We now propose the heap structure-based pilot assignment strategy. To do this, the following definition is invoked.

Definition 1: Min heap $(\mathcal{H}^{\min})$ is a tree-based structure, where $\mathcal{H}_{\mathtt{p}}^{\min}$ is a parent node of an arbitrary node $\mathcal{H}_{\mathtt{c}}^{\min}$. Then, the key of $\mathcal{H}_{\mathtt{p}}^{\min}$ is less than or equal to that of $\mathcal{H}_{\mathtt{c}}^{\min}$. In a max heap $(\mathcal{H}^{\max})$, the key of $\mathcal{H}_{\mathtt{p}}^{\max}$ is greater than or equal to that of $\mathcal{H}_{\mathtt{c}}^{\max}$ [45]. We note that a node of heap contains not only the keys (values) to build a heap structure, but also other specific parameters depending on storage purposes. Therefore, if a node is moved on the heap, all its parameters are also moved in company.

Let $\mathcal{H} \in \{\mathcal{H}^{\min}, \mathcal{H}^{\max}\}$, the following main operations are involved:

- **Generate a heap** $\big(\mathcal{G}(\mathbf{x}, \{\mathbf{y}\}) \rightarrowtail \mathcal{H}\big)$: The size and keys of $\mathcal{H}$ follow the size and values of vector $\mathbf{x}$, where $\{\mathbf{y}\}$ is a set of parameters.
- **Find min/max value** $\big(\mathcal{H} \rightarrow (x, \{\mathbf{y}\})\big)$: To return the root key x and the parameter set $\{\mathbf{y}\}$ of $\mathcal{H}$.
- **Extract the root node** $\big(\mathcal{H} \vdash (x, \{\mathbf{y}\})\big)$: To pop the root node out of $\mathcal{H}$ (i.e., extract the max/min value), and then assign to $(x, \{\mathbf{y}\})$. Next, $\mathcal{H}$ is updated to restore the heap condition.
- **Replace and Sift-down** $\big(\mathcal{H} \dashv (x, \{\mathbf{y}\})\big)$: To replace the root node with the key x and its parameter set $\{\mathbf{y}\}$, and then, move the root node down in the tree to restore the heap condition.

Let $\mathbf{u}_i^{\mathtt{ID}}$ be the i-th column of the identity matrix of size τ ($\mathbf{I}_\tau$). Then, the feasible set of the j-th column variable of matrix $\boldsymbol{\Upsilon}$, corresponding to UE j, is determined as $\boldsymbol{\upsilon}_j \in \{\mathbf{u}_i^{\mathtt{ID}} | i \in \mathcal{T}_{\mathtt{p}}\}$, $j \in \mathcal{T}_{\mathtt{u}}$, satisfying constraint (50b). From Theorem 4, if the i-th pilot is assigned to two arbitrary UEs $j, k \in \mathcal{T}_{\mathtt{u}}$, we have

$$\sum_{j' \in \mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} \boldsymbol{\upsilon}_j^H \boldsymbol{\upsilon}_{j'} = \sum_{j' \in \mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} \boldsymbol{\upsilon}_k^H \boldsymbol{\upsilon}_{j'} = \sum_{j' \in \mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} (\mathbf{u}_i^{\mathtt{ID}})^H \boldsymbol{\upsilon}_{j'}.$$

We define a pilot-reused coefficient (PRC) of the i-th pilot by $\bar{\beta}_i \triangleq \sum_{j' \in \mathcal{T}_{\mathtt{u}}} \tilde{\beta}_{j'} (\mathbf{u}_i^{\mathtt{ID}})^H \boldsymbol{\upsilon}_{j'}$, and rewrite (51) equivalently as

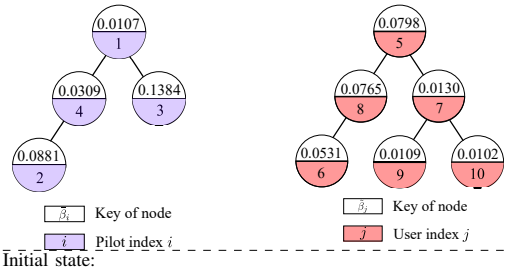
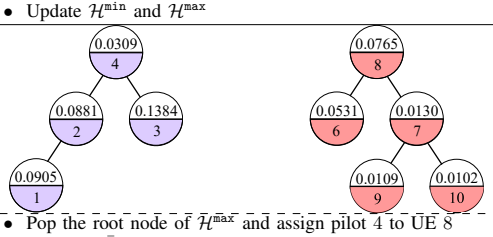
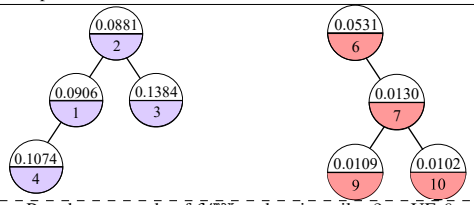
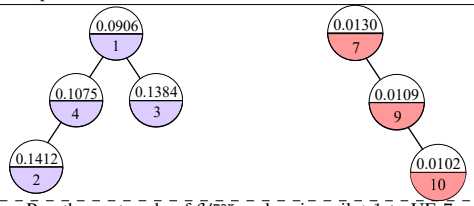
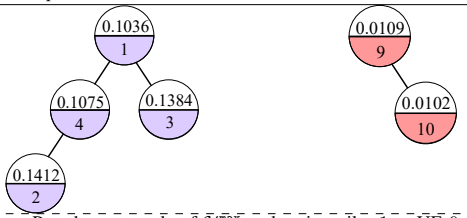
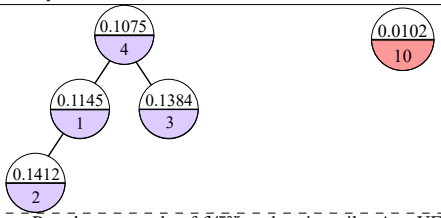
$$\min_{\boldsymbol{\Upsilon}} \quad \max_{i \in \mathcal{T}_{\mathtt{p}}} \{\bar{\beta}_i\}, \tag{52a}$$

$$\text{s.t.} \quad \boldsymbol{\upsilon}_j \in \{\mathbf{u}_i^{\mathtt{ID}} | i \in \mathcal{T}_{\mathtt{p}}\},\ j \in \mathcal{T}_{\mathtt{u}}. \tag{52b}$$

It is realized that if the i-th pilot is assigned to UE j, the PRC of the i-th pilot increases by a factor of $\tilde{\beta}_j$. To minimize the maximum of PRCs, a heuristic assignment is executed such that the pilot with the smallest PRC is assigned to UE j with the largest $\tilde{\beta}_j$. The following example is to illustrate the procedure of the proposed heap-based pilot assignment.

Example: we consider a scenario where $\tau = 4$ pilots need to be assigned to $U = 10$ UEs with a given large-scale fading as: $[\tilde{\beta}_1, \cdots, \tilde{\beta}_{10}] = [0.0107, 0.0881, 0.1384, 0.0309, 0.0798, 0.0531, 0.0130, 0.0765, 0.0109, 0.0102]$. The assignment progress is described in Table II.

TABLE II
STATES OF HEAP STRUCTURES PER ITERATION

#Iter.	Min-heap $(\mathcal{H}^{\min})$	Max-heap $(\mathcal{H}^{\max})$
0		
	Initial state: • τ pilots are assigned to the first τ UEs, leading to $\mathcal{H}^{\min}$ with τ nodes • $U - \tau$ remaining UEs are put into $\mathcal{H}^{\max}$ Execution: • Pop the root node of $\mathcal{H}^{\max}$, and assign pilot 1 to UE 5 • Compute $\bar{\beta}_1 = 0.0107 + 0.0798 = 0.0905$ • Update $\mathcal{H}^{\min}$ and $\mathcal{H}^{\max}$	
1		
	• Pop the root node of $\mathcal{H}^{\max}$ and assign pilot 4 to UE 8 • Compute $\bar{\beta}_4 = 0.0309 + 0.0765 = 0.1074$ • Update $\mathcal{H}^{\min}$ and $\mathcal{H}^{\max}$	
2		
	• Pop the root node of $\mathcal{H}^{\max}$ and assign pilot 2 to UE 6 • Compute $\bar{\beta}_2 = 0.0881 + 0.0531 = 0.1412$ • Update $\mathcal{H}^{\min}$ and $\mathcal{H}^{\max}$	
3		
	• Pop the root node of $\mathcal{H}^{\max}$ and assign pilot 1 to UE 7 • Compute $\bar{\beta}_1 = 0.0906 + 0.0130 = 0.1036$ • Update $\mathcal{H}^{\min}$ and $\mathcal{H}^{\max}$	
4		
	• Pop the root node of $\mathcal{H}^{\max}$ and assign pilot 1 to UE 9 • Compute $\bar{\beta}_1 = 0.1036 + 0.0109 = 0.1145$ • Update $\mathcal{H}^{\min}$ and $\mathcal{H}^{\max}$	
5		
	• Pop the root node of $\mathcal{H}^{\max}$ and assign pilot 4 to UE 10 which complete the pilot assignment	

Algorithm 2 Proposed Heap-Based Pilot Assignment for MSE Minimization Problem (51)

1: Compute $\tilde{\boldsymbol{\beta}} \triangleq [\tilde{\beta}_j]_{j\in\mathcal{T}_{\mathtt{u}}}$ as in **Theorem** 4.
2: Randomly assign τ pilots to the first τ UEs in $\mathcal{T}_{\mathtt{u}}$, yielding $\boldsymbol{\upsilon}_j,\ \forall j = 1, \cdots, \tau$.
3: Execute $\mathcal{G}([\tilde{\boldsymbol{\beta}}]_{1:\tau}, \{\boldsymbol{\upsilon}_j\}_{j=1,\cdots,\tau}) \rightarrowtail \mathcal{H}^{\mathtt{min}}$.
4: Execute $\mathcal{G}([\tilde{\boldsymbol{\beta}}]_{\tau+1:U}, \{\tau+1, \cdots, U\}) \rightarrowtail \mathcal{H}^{\mathtt{max}}$.
5: **while** $\mathcal{H}^{\mathtt{max}} \neq \emptyset$ **do**
6: $\quad \mathcal{H}^{\mathtt{max}} \vdash (\tilde{\beta}_{j'}, \{j'\})$. {*Root node is removed from* $\mathcal{H}^{\mathtt{max}}$}
7: $\quad \mathcal{H}^{\mathtt{min}} \rightarrow (\tilde{\beta}_i, \{\boldsymbol{\upsilon}_i\})$.
8: $\quad \boldsymbol{\upsilon}_{j'} := \boldsymbol{\upsilon}_i$.
9: $\quad \mathcal{H}^{\mathtt{min}} \dashv (\tilde{\beta}_i + \tilde{\beta}_{j'}, \{\boldsymbol{\upsilon}_i\})$.
10: **end while**
11: Concatenate assignment variable vectors as $\boldsymbol{\Upsilon} := [\boldsymbol{\upsilon}_1, \cdots, \boldsymbol{\upsilon}_U]$.
12: **Output:** Pilot assignment matrix $\bar{\boldsymbol{\Xi}} = \boldsymbol{\Xi}\boldsymbol{\Upsilon}$.

The proposed algorithm for pilot assignment is summarized in Algorithm 2. It takes the complexity of $\mathcal{O}\big(U\log_2(\tau U)\big)$ for deriving the assignment solution, which is relatively low complexity. For simplicity, this training strategy is referred to as Heap-FD, in which Algorithm 2 is operated twice, i.e., with $U = L$ for UL channel estimation in the first phase, and with $U = K$ for achieving DL and CCI channel estimates in the second phase. On the other hand, the training strategy for HD systems can be done by setting $U = K + L$, called Heap-HD. For a given τ-length of pilot sequences, Heap-FD requires the training time of 2τ, while Heap-HD needs only τ. However, it is anticipated that the channel estimate quality of Heap-FD would be better than Heap-HD, due to smaller number of UEs sharing the same pilot set in Heap-FD.

Remark *7:* The SE-EE problem (10) can be reformulated as a worst-case robust design by treating CSI errors as noise. More specifically, we use channel estimates (rather than the perfect ones) to perform data transmission. The additional component introduced by CSI errors in the denominators of SINRs is a linear function, and thus, can be easily tackled by our proposed methods given in Sections IV and V. We refer the interested reader to [17, Sec. V] for further details of the derivations.

VII. Numerical Results

In this section, we provide numerical examples to quantitatively evaluate the performance of the proposed FD CF-mMIMO.

A. Simulation Setup and Parameters

A system topology illustrated in Fig. 2 is considered, where all APs and UEs are located within a circle of 1-km radius. The entries of the fading loop channel $\mathbf{G}^{\mathtt{SI}}_{mm}, \forall m \in \mathcal{M}$ are modeled as independent and identically distributed Rician RVs, with the Rician factor of 5 dB [19]. The large-scale fading of other channels is modeled as [23]

$$\beta = 10^{\frac{\mathtt{PL}(d)+\sigma_{\mathtt{sh}}z}{10}}, \tag{53}$$

where $\beta \in \{\beta^{\mathtt{AA}}_{mm'}, \beta^{\mathtt{d}}_{km}, \beta^{\mathtt{u}}_{m\ell}, \beta^{\mathtt{cci}}_{k\ell}\}, \forall m, m' \in \mathcal{M}, k \in \mathcal{K}, \ell \in \mathcal{L}$ and $m \neq m'$; The shadow fading is considered as an RV $z \in \{z^{\mathtt{AA}}_{mm'}, z^{\mathtt{d}}_{km}, z^{\mathtt{u}}_{m\ell}, z^{\mathtt{cci}}_{k\ell}\} \sim \mathcal{N}(0,1)$ with standard deviation $\sigma_{\mathtt{sh}} = 8$ dB. The three-slope model for the path loss in dB is given by [23], [47]

$$\mathtt{PL}(d) = -140.7 - 35\log_{10}(d) + 20c_0\log_{10}(\frac{d}{d_0}) + 15c_1\log_{10}(\frac{d}{d_1}), \tag{54}$$

where $d \in \{d^{\mathtt{AA}}_{mm'}, d^{\mathtt{d}}_{km}, d^{\mathtt{u}}_{m\ell}, d^{\mathtt{cci}}_{k\ell}\}$ is the distance between transceivers as corresponding to β; d_i, with $i = \{0, 1\}$, denotes the reference distance and $c_i \triangleq \max\{0, \frac{d_i - d}{|d_i - d|}\}$. Note that the distances d and d_i in (54) are measured in km. Unless specifically stated otherwise, other parameters are given in Table III, where all APs are assumed to have the same power budget $P^{\mathtt{max}}_{\mathtt{AP}} = P^{\mathtt{max}}_{\mathtt{AP}_m}, \forall m$. Herein, the parameters of power consumption and PA efficiencies follow the study in [41]. We use the modeling tool YALMIP in the MATLAB environment. The SEs are divided by $\ln 2$ to be presented in bits/s/Hz.

For comparison, the following two known schemes are considered:

1) "Co-mMIMO:" A BS is deployed at the center of the considered area to serve all UEs. To conduct a fair comparison, the centered-BS is equipped with $N = \sum_{m\in\mathcal{M}} N_m$ number of antennas with the total power budget of $MP^{\mathtt{max}}_{\mathtt{AP}}$ = 43 dBm.
2) "SC-MIMO:" Under the same setup with CF-mMIMO, each UE is only served by one AP, but each AP can serve more than one UE. To make ZF feasible, the number of

TABLE III
Simulation Parameters

Parameter	Value
System bandwidth, B	10 MHz
Reference distances, (d_0, d_1)	(10, 50) m
Residual SiS, $\rho^{\mathtt{RSI}} = \rho^{\mathtt{RSI}}_{mm},\ \forall m$	-110 dB [46]
Noise power at receivers	-104 dBm
Number of APs and UEs, $(M,\ K,\ L)$	(64, 10, 10)
Number of antennas per AP, $N_m, \forall m$	2
Rate threshold, $\bar{R} = \bar{R}^{\mathtt{d}}_k = \bar{R}^{\mathtt{u}}_\ell,\ \forall k, \ell$	0.5 bits/s/Hz
PA efficiency at $\mathtt{AP}_m,\ \nu_{\mathtt{AP}_m}, \forall m$	0.39
PA efficiency at $\mathtt{U}^{\mathtt{u}}_\ell,\ \nu^{\mathtt{u}}_\ell,\ \forall \ell$	0.3
Backhaul traffic power, $P^{\mathtt{bh}}$	0.25 W/(Gbits/s)
Baseband power, $P^{\mathtt{d}}_{km} = P^{\mathtt{u}}_m, \forall k, m$	0.1 W
APs' power in active/sleep modes, $(P^{\mathtt{a}}_{\mathtt{AP}_m}, P^{\mathtt{s}}_{\mathtt{AP}_m}), \forall m$	(10, 2) W
APs' circuit operation power, $P^{\mathtt{cir}}_{\mathtt{AP}_m}, \forall m$	1 W
UEs' circuit operation power, $P^{\mathtt{d,cir}}_k = P^{\mathtt{u,cir}}_\ell, \forall k, \ell$	0.1 W
Power budget at UL UEs, $P^{\mathtt{max}}_\ell, \forall \ell$	23 dBm
Total power budget for all APs, $MP^{\mathtt{max}}_{\mathtt{AP}}$	43 dBm

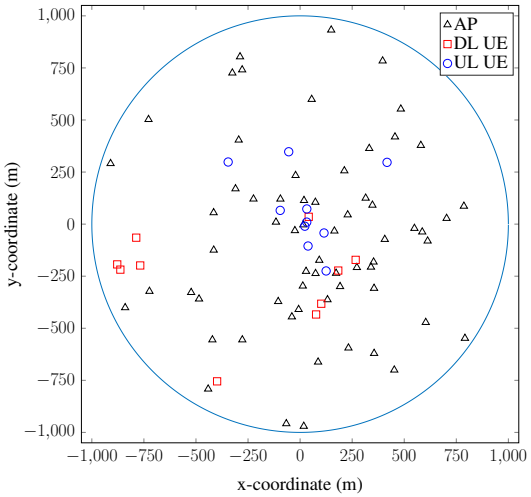


Fig. 2. A system topology with $M = 64$ and $K = L = 10$ located within a circle of 1-km radius is used in numerical examples.

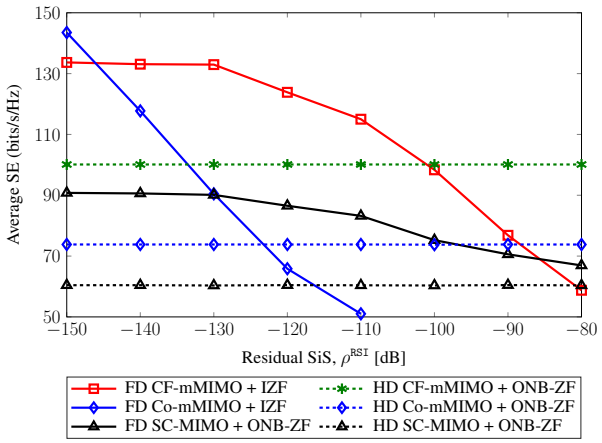


Fig. 3. Average SE versus the residual SiS, ρ^{RSI}, for different schemes with both FD and HD operations.

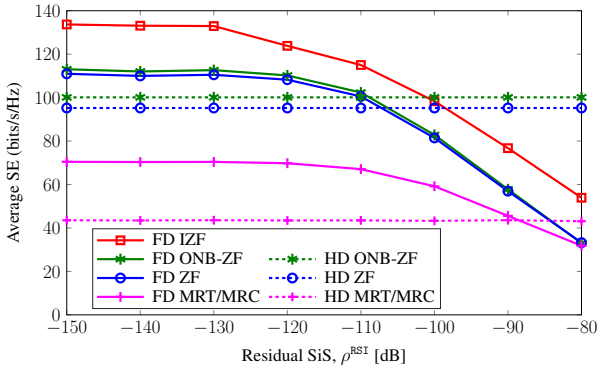


Fig. 4. Average SE versus the residual SiS, ρ^{RSI}, for different transmission strategies in CF-mMIMO.

UEs served by one AP must be less than its own number of antennas.

Both FD and HD operations are employed to evaluate the performance of those schemes. For HD operation, DL and UL transmissions are separately carried out in two independent communication time blocks. As a result, there is no CCI at DL UEs, and no RSI and IAI on UL reception, but the achieved SE and EE are devided by two. To show the effectiveness of the proposed ZF- and IZF-based transmissions presented in Sections IV and V, respectively, we additionally examine the following transmission strategies:

1) "ONB-ZF:" The precoder matrix for DL transmission $\mathbf{W}^{\mathrm{ONB\text{-}ZF}}$ is computed as in (41), while UL reception adopts the ZF-SIC receiver similar to IZF.
2) "MRT/MRC:" MRT and MRC are applied to DL and UL, respectively. This is easily done by replacing $\mathbf{H}^{\mathrm{ZF}}$ and $\mathbf{A}^{\mathrm{ZF}}$ with $(\mathbf{H}^{\mathrm{d}})^H$ and $(\mathbf{H}^{\mathrm{u}})^H$, respectively.

B. Numerical Results for SE Performance

Fig. 3 depicts the average SE performance as a function of the residual SiS ρ^{RSI} for different schemes in both FD and HD operations. We recall that the SI has no effect on the performance of HD-based schemes. The first observation is that FD-based schemes outperform HD counterparts at a sufficiently small level of ρ^{RSI}. In particular, at $\rho^{\mathrm{RSI}} = -130$ dB, FD CF-mMIMO and FD SC-MIMO provide more than 40% SE performance gain over those of HD ONB-ZF designs.

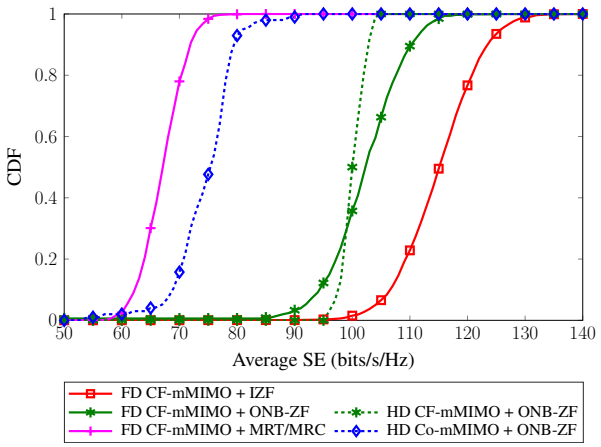


Fig. 5. Cumulative distribution function versus the average SE.

The SE of FD Co-mMIMO scheme and its HD counterpart confirms the promising performance of the FD system. Specifically, when the residual SiS is very small (i.e., ρ^{RSI} = -150 dB), the performance of FD Co-mMIMO nearly doubles that of HD Co-mMIMO. However, its performance is dramatically degraded when ρ^{RSI} increases. The reason is that the FD Co-mMIMO system with one centered BS serving all UEs in a large area must allocate a high transmit power to far DL UEs, leading to higher SI power. On the other hand, when ρ^{RSI} is more severe, the probabilities of infeasibility of the FD schemes are higher, resulting in the degraded performance. Another interesting observation is that the proposed FD CF-mMIMO scheme provides the best performance among FD-based schemes and significantly better performance than HD for a wide range of ρ^{RSI}, including the practical value of ρ^{RSI} = -110 dB [46]. These results further confirm that FD operation is well suited for CF-mMIMO systems.

To evaluate the effectiveness of the proposed ZF and IZF transmission strategies in CF-mMIMO, we compare the average SE of our designs with two simple transmission designs: ONB-ZF and MRT/MRC, as shown in Fig. 4. With our proposed FD IZF design, the SE improves significantly. The ONB-ZF-PCA precoding in IZF not only cancels MUI for DL transmission, but also depresses the effect of residual SI on UL reception. In addition, ONB-ZF provides slightly better performance than ZF in both FD and HD. However, the gap between FD ONB-ZF- and FD ZF-based designs is gradually small when ρ^{RSI} increases. It reflects the fact that the effect of residual SI on the system performance of CF-mMIMO is more serious than that of MUI, which brings less benefit of using SIC in the MUI cancellation. The MRT/MRC design is always inferior in both FD and HD operations since it is unable to manage the network interference effectively. The difficulty in handling the IAI and residual SI of ONB-ZF and ZF designs causes their FD system to provide worse performance than the proposed FD IZF.

Fig. 5 shows the cumulative distribution function (CDF) of different transmission strategies in FD CF-mMIMO with the references to HD CF-mMIMO and HD Co-mMIMO. The figure clearly shows that the feasibility probabilities of all the considered schemes are smaller when the SE is higher. As expected, FD CF-mMIMO with the proposed IZF and ONB-

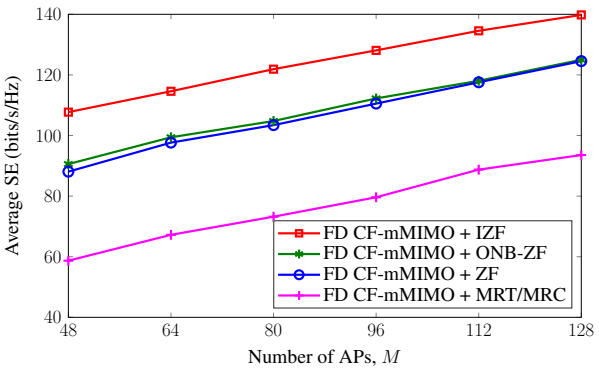


Fig. 6. Average SE versus the number of APs with different transmission strategies in FD CF-mMIMO.

ZF designs outperforms others. In addition, FD CF-mMIMO with IZF offsets SE with about 12 bits/s/Hz more than FD CF-mMIMO with ONB-ZF. Although the ONB-ZF design is able to cancel MUI, it still suffers the large amount of IAI, which is even more severe in FD CF-mMIMO with the dense AP deployment. Its performance is therefore slightly better than that of HD CF-mMIMO at mid-point and 95-percentile point. FD CF-mMIMO with MRT/MRC provides worst performance, which can be explained as follows. MRT/MRC for DL/UL transmission using the channel conjugates (i.e., $(\mathbf{H}^{\mathrm{d}})^H$ and $(\mathbf{H}^{\mathrm{u}})^H$) is inapplicable to handle the residual SI and also totally passive with the IAI. Such interference is inherent to the MRT/MRC design in FD CF-mMIMO. The CDFs with respect to the SE again validate the advantage of IZF with ONB-ZF for MUI cancellation and PCA procedure for IAI and residual SI depression.

In Fig. 6, we plot the average SE of different transmission strategies in FD CF-mMIMO versus the number of APs, $M \in [48, 128]$. It is straightforward to see that increasing the number of APs causes stronger IAI in CF-mMIMO networks. However, as can be seen from the figure, the SEs of all the considered transmission strategies are monotonically improved when M increases. This results imply that it is enough for each AP to select a suitable number of UEs around it, without causing much interference to other APs.

C. Numerical Results for EE Performance

Before providing the EE performance, we show the status of APs (i.e., active and sleep modes) in Fig. 7 and operation behavior of APs (i.e., DL/UL or FD) in Fig. 8, gaining more insights into the effect of AP selection. We consider the system topology illustrated in Fig. 2, and the IZF design to maximize the EE. It can be seen in Fig. 7 that most APs located far way from UEs switch off to reduce power consumption, since the large effect of path loss makes power consumption for these APs inefficient. In contrast, APs in the area of dense UEs become strongly active in order to enhance the SE $\tilde{F}_{\mathrm{SE}}(\mathbf{\Lambda}_{\mathrm{d}}, \mathbf{\Lambda}_{\mathrm{u}})$, which dominates the loss caused by ϕ in (35). This result suggests an interesting observation that each UE should be served by a small subset of active APs to manage the network interference more effectively. Fig. 8 shows how the power budget at an active AP is allocated to UEs. Intuitively, the active APs dynamically connect to near UEs for the EE improvement, and thus, HD (DL/UL) or FD mode is selected

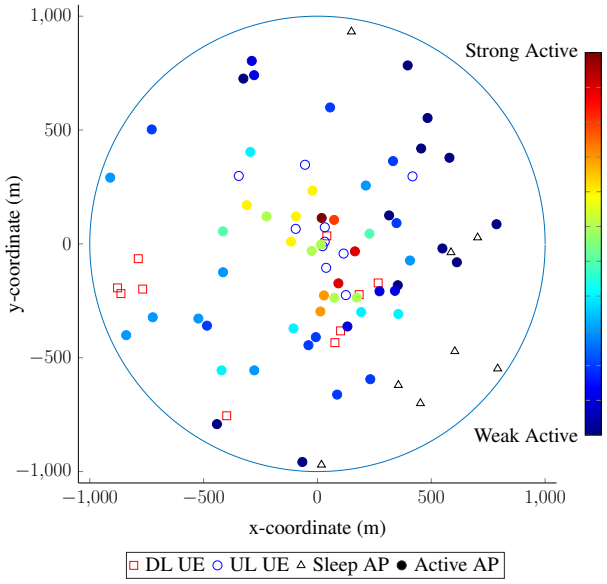


Fig. 7. Active and sleep modes of APs using the system topology in Fig. 2.

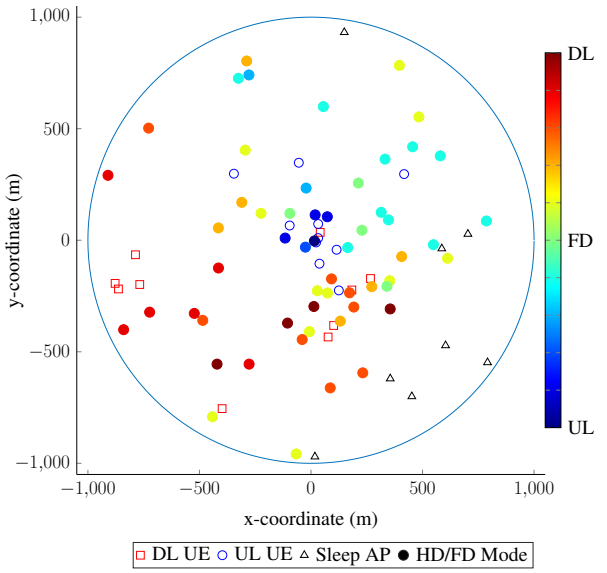


Fig. 8. Operation behavior of APs using the system topology in Fig. 2

depending on either DL or UL transmission of near UEs, as long as their rate thresholds are satisfied. Specifically, APs are operated in DL and UL modes when they are located close to DL and UL UEs, respectively, and the FD mode, otherwise. In other words, the proposed method allows active APs to dynamically switch between DL/UL (in HD) and FD modes based on the channel conditions of all UEs to alleviate the IAI, residual SI and CCI. These observations reflect the importance of the joint design of AP-UE association and AP selection to obtain the maximum EE performance. It is expected that FD CF-mMIMO with AP selection outperforms the case without AP selection, as discussed in the following part.

We now examine the EE performance versus the number of UEs ($K = L$) in FD CF-mMIMO with different transmission strategies, as illustrated in Fig. 9. To ensure a high feasibility of the considered schemes even when the number of UEs is large, we set the number of APs to $M = 256$. The EE of the considered transmission strategies first increases, approaches the optimal point, and then decreases, as the number of UEs

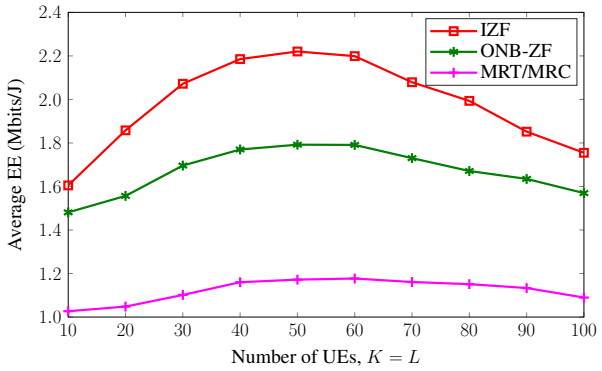


Fig. 9. Average EE versus the number of UEs in FD CF-mMIMO with $M = 256$ and $\bar{R} = 0.5$ bits/s/Hz.

increases. This phenomenon is attributed to the fact that, for small and medium numbers of UEs, the SE improvement dominates the total power consumption, leading to the significantly enhanced EE performance. However, when the number of UEs is very large (i.e., $K = L > 50$), the SE is slightly improved, even reduced, due to stronger network interference, while the total power consumption increases quickly since more APs are required to be active to serve larger number of UEs. Nevertheless, our proposed IZF design outperforms the baseline ones.

Fig. 10 plots the average EE as a function of the number of APs M for $\bar{R} \in \{0, 0.5\}$ bits/s/Hz, with and without AP selections. For benchmarking purpose, we consider two baseline schemes: (i) Power consumption of sleep APs is completely set to zero, named as "IZF with perfect AP selection (IZF w/ Perf. AP Sel.);" (ii) AP-UE association and AP selection are used in company with the channel matrix transpose and an equal weight coefficient for all UEs, referred to as "AP selection without weight beamforming design (AP Sel. w/o WBF)." It can be seen that the IZF design with AP selection obtains much better EE performance than without AP selection, i.e., up to about 50% EE gain. However, the EE with AP selection degrades when the number of APs becomes large, since the sleep APs still consume a fixed power in sleep modes, as given in (7). This also explains why the perfect AP selection achieves the best performance among all the schemes and gradually increases along with the number of APs. In addition, increasing M brings no benefit to the schemes without AP selection. We can also see that the performance gaps between $\bar{R} = 0$ bits/s/Hz and $\bar{R} = 0.5$ bits/s/Hz of all considered schemes are narrower when M increases. The larger the number of APs, the higher the probability for UEs to efficiently select a subset of APs with good channel conditions, leading to better rate fairness among all UEs.

D. Numerical Results for Heap-Based Pilot Assignment Algorithm 2

It can be easily foreseen that the quality of channel estimates mainly depends on the relationship between the number of UEs and dimension of pilot set (or pilot length, τ). To evaluate the performance of the proposed FD training strategy, we first investigate the normalized MSE (NMSE) as a function of the number of UEs. As depicted in Fig. 11, we consider four

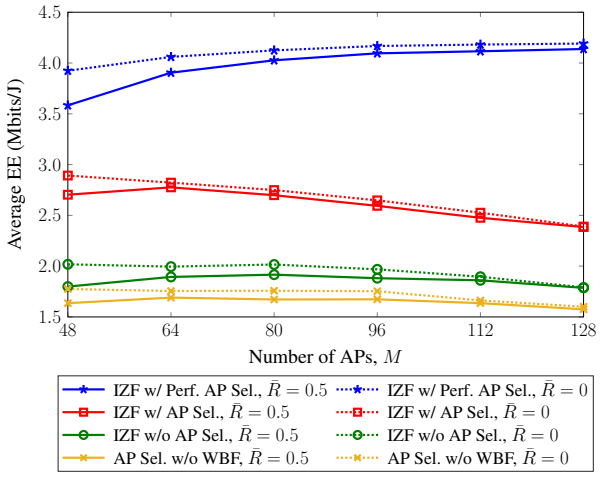


Fig. 10. Average EE versus the number of APs in FD CF-mMIMO under different types of AP selections, with $\bar{R} \in \{0, 0.5\}$ bits/s/Hz.

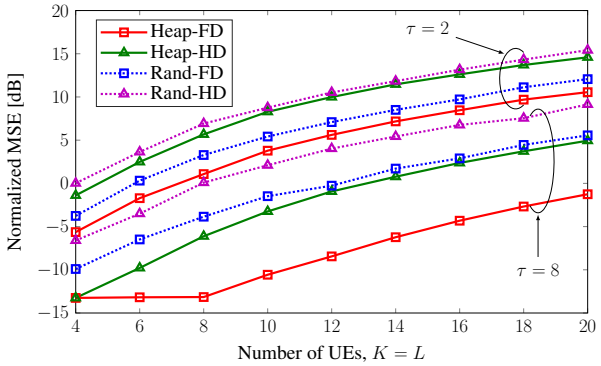


Fig. 11. Normalized MSE versus the number of UEs in CF-mMIMO with $\tau \in \{2, 8\}$.

strategies: two heap structures for pilot assignment (Heap-FD and Heap-HD), and two random pilot assignments (Rand-FD and Rand-HD). As expected, the proposed heap training schemes outperform the random ones. It can also be observed that FD training strategies offer better performance in terms of NMSE compared to HD ones, by exploiting larger dimension of pilot sequences more efficiently. In particular, when $K = L > \tau$, NMSE of the proposed Heap-FD is around 5 dB and 7 dB less than Heap-HD, corresponding to $\tau = 2$ and $\tau = 8$, respectively. However, we note that the FD training strategy requires a double training time over its HD counterpart, leading to the difference of the effective time for data transmission. The SE under imperfect CSI can be expressed as

$$\hat{F}_{\mathrm{SE}}(\mathbf{w}, \mathbf{p}, \boldsymbol{\alpha}) = \frac{\tau_{\mathrm{c}} - \tau_{\mathrm{t}}}{\tau_{\mathrm{c}}} F_{\mathrm{SE}}(\mathbf{w}, \mathbf{p}, \boldsymbol{\alpha}), \tag{55}$$

where τ_{c} and τ_{t} are the coherent time and training time, respectively. We now plot the SE performance for the worst-case robust design by taking into account the channel estimation. In Fig 12, we set $\tau_{\mathrm{c}} = 200$, $\tau_{\mathrm{t}} = 2\tau$ for FD and $\tau_{\mathrm{t}} = \tau$ for HD. Unsurprisingly, Heap-FD schemes outperform HD ones, and their performance gaps are even more remarkable when the number of UEs increases. This again demonstrates the effectiveness of the proposed Heap-based pilot assignment algorithm for FD CF-mMIMO by reaping both the advantages of higher dimension of pilot sequences for training and FD for data transmission.

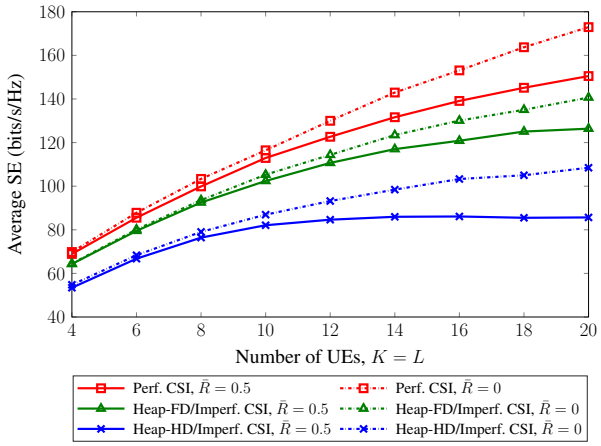
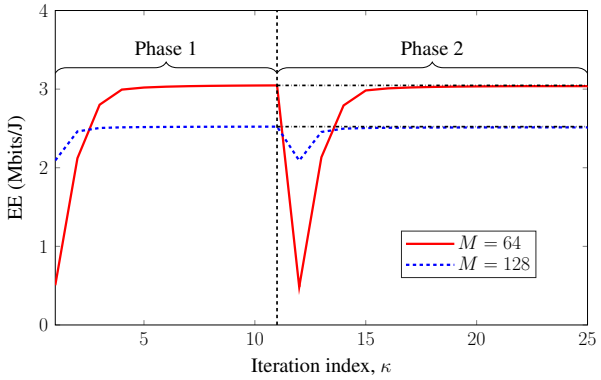
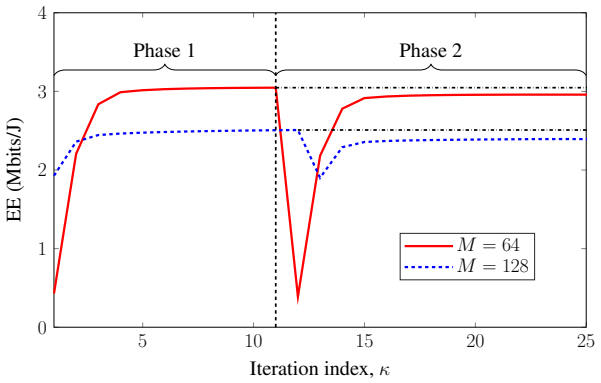


Fig. 12. Average SE versus the number of UEs in CF-mMIMO with $\tau = 8$ and $\bar{R} \in \{0, 0.5\}$ bits/s/Hz.



(a) EE convergence behavior of Algorithm 1 in FD CF-mMIMO, with the per-AP power signal ratio $\varpi = 0.1\%$ of $1/M$.



(b) EE convergence behavior of Algorithm 1 in FD CF-mMIMO, with the per-AP power signal ratio $\varpi = 1\%$ of $1/M$.

Fig. 13. Typical EE convergence behaviors of Algorithm 1 over a random channel realization.

E. Convergence Behavior and Computational Capability of Algorithm 1

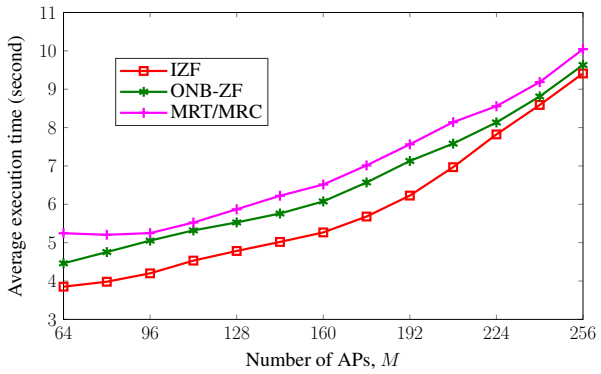
We have numerically observed that the convergence behavior of the SE optimization in Algorithm 1 is similar to that of the EE. For the sake of brevity, only the EE convergence performance is plotted in Fig. 13 by setting $\eta = 0$, with $M \in \{64, 128\}$. Since each AP has an average percentage of rate contributed to UEs $\hat{\varpi} = 1/M$, we consider ϖ given in (16) as 0.1% and 1% of $1/M$, as illustrated in Fig. 13(a) and Fig. 13(b), respectively. As seen, Algorithm 1 converges very fast, and attains 99% EE performance within about 10 iterations for both the first phase (Steps 1-9 in Algorithm 1) and second phase (Step 11 in Algorithm 1). The figure also clearly demonstrates the effect of the selection of per-AP power signal ratio ϖ on the EE performance. For $\varpi = 0.1\%$ of $1/M$, the gap between two phases is very small, i.e., lower than 0.05% of the EE achieved in phase 1. For $\varpi = 1\%$ of $1/M$, the EE loss in phase 2 increases up to 3% and 5% , corresponding to M = 64 and M =128, respectively. It simply implies that the value of ϖ should be properly chosen to not only achieve a good performance, but also recover an exact binary value of $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$.

Finally, we provide the average execution time of the proposed designs, which mainly aims at showing how their computational complexities scales with the network size. The codes are implemented in MATLAB with the modeling toolbox YALMIP running on a computer of Intel(R) Core(TM) i7-6700 CPU @ 3.4 GHz, RAM 16 GB and Windows 10. Fig. 14(a) shows that the execution time of all the proposed designs slightly increases even when M increases rapidly, since the problem is less dependent on M. However, the execution time scales exponentially with respect to the number of UEs, as shown in Fig. 14(b). Although the computational complexities of IZF, ONB-ZF, and MRT/MRC design are similar, the distinct feasible regions would lead to different processing times. The higher execution time of MRT/MRC is due to its smaller feasible region. To be more comprehensive, Fig. 14(c) plots the normalized effective sum power of IAI and RSI (i.e., $\|\tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{W}\|^2/\sigma^2$), and the execution time versus the percentage of $\bar{N}$ eigenvalues in (39) over the total eigenvalues, $\delta \in (0.6, 0.99]$. Clearly, the normalized sum power of IAI and RSI for ONB-ZF and MRT/MRC are unchanged, regardless of the value of δ. The reason is that ONB-ZF has to preserve the structure of the ZF matrix, while MRT/MRC is based on the transpose of channel responses to compute the precoder/receiver. More importantly, the IZF transmission design is capable of providing lower *leakage* IAI/RSI power and execution time at the large value of δ.

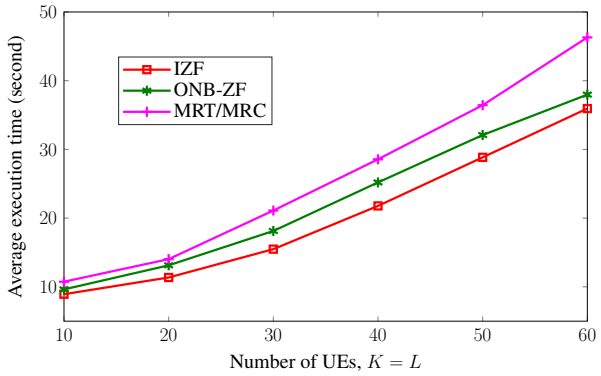
VIII. Conclusion

We have investigated the SE and EE of an FD CF-mMIMO network by jointly optimizing power control, AP-UE association and AP selection. The realistic power consumption model, which accounts for data transmission, baseband processing and circuit operation, has been taken into consideration in characterizing the EE performance. Also, the special relationship between binary and continuous variables has been efficiently exploited to reduce the number of optimization variables. First, we have derived the iterative procedure based on the ICA framework and Dinkelbach method to solve the ZF-based problem, where each iteration only solves a simple convex program. Aiming at efficient network interference management, we have then proposed an improved ZF-based transmission by incorporating ONB-and-PCA in the DL, and SIC in the UL. In addition, a novel and low-complexity pilot assignment algorithm based on the heap structure has been developed to improve the quality of channel estimates.

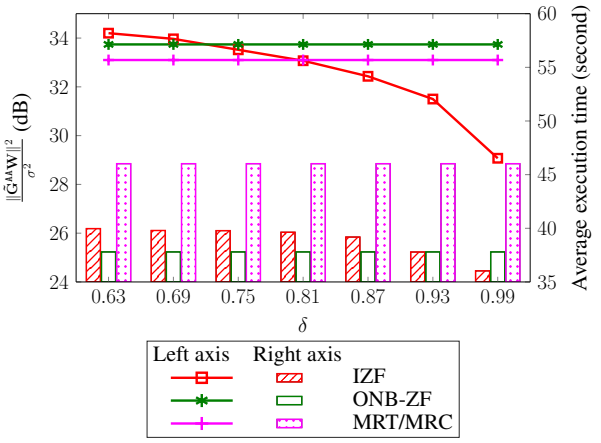
The proposed algorithm admitted fast convergence rate,



(a) Average execution time versus the number of APs, M.



(b) Average execution time versus the number of UEs, with $M = 256$.



(c) Normalized sum power of IAI and RSI and execution time versus δ, with $M = 256$, $K = L = 60$, and $\mathbf{W} \in \{\mathbf{W}^{\mathtt{IZF}}, \mathbf{W}^{\mathtt{ONB-ZF}}, (\mathbf{H}^{\mathtt{d}})^H\}$.

Fig. 14. Average execution time of Algorithm 1 for different transmission strategies in FD CF-mMIMO.

and showed to significantly outperform SC-MIMO and Co-mMIMO in terms of SE and EE by jointly optimizing the parameters involved. Via presented results, it can be concluded that FD CF-mMIMO with IZF transmission design is much more robust against the effects of residual SiS and IAI and requires lower execution time than ZF, ONB-ZF and MRT/MRC. Numerical results also showed that much better EE performance can be yielded by our joint design together with the AP selection.

Appendix A: Proof of Lemma 1

Suppose that the optimal solution for (10) is found as a 2-tuple $\mathbf{u}^* \triangleq (\mathbf{s}^*, \boldsymbol{\mu}^*)$, where $\mathbf{s}^* \triangleq (\mathbf{w}^*, \mathbf{p}^*, \boldsymbol{\alpha}^* | \ \alpha^*_{km} = 0 \ \& \ \mathbf{w}^*_{km} \neq \mathbf{0})$. Let $\tilde{\mathbf{s}} \triangleq (\tilde{\mathbf{w}}, \mathbf{p}^*, \boldsymbol{\alpha}^* | \alpha^*_{km} = 0 \ \& \ \tilde{\mathbf{w}}_{km} = \mathbf{0})$, yielding $\tilde{\mathbf{u}} = (\tilde{\mathbf{s}}, \boldsymbol{\mu}^*)$. We are now in a position to prove $F(\mathbf{u}^*) \leq F(\tilde{\mathbf{u}})$ as $\tilde{\mathbf{u}}$ is an optimal point. Inspired from [48], it can be realized that the numerator and denominator of $\gamma^{\mathtt{d}}_k(\mathbf{s}^*)$ in (2) remain unchanged for any $\mathbf{w}^*_{km}$. From (8), $F_{\mathtt{SE}}(\mathbf{s}^*)$ is the same with respect to $\{\mathbf{w}^*_{km}\}$. Moreover, α_{km} coupled with $\mathbf{w}_{km}$ in (4) gives $\gamma^{\mathtt{u}}_\ell(\mathbf{s}^*) = \gamma^{\mathtt{u}}_\ell(\tilde{\mathbf{s}})$, and thus, $F_{\mathtt{SE}}(\mathbf{s}^*) = F_{\mathtt{SE}}(\tilde{\mathbf{s}})$. On the other hand, if $\|\mathbf{w}^*_{km}\|^2 > 0$, then $P_{\mathtt{D}}(\mathbf{u}^*) \geq P_{\mathtt{D}}(\tilde{\mathbf{u}})$ given by the first term in (6). The equality between $P_{\mathtt{D}}(\mathbf{u}^*)$ and $P_{\mathtt{D}}(\tilde{\mathbf{u}})$ holds if $\mu^*_m = 0$. As a result, $P_{\mathtt{T}}(\mathbf{u}^*) \geq P_{\mathtt{T}}(\tilde{\mathbf{u}})$ leads to $F_{\mathtt{EE}}(\mathbf{u}^*) \leq F_{\mathtt{EE}}(\tilde{\mathbf{u}})$ as well as $F(\mathbf{u}^*) \leq F(\tilde{\mathbf{u}})$. Moreover, the indexes k and m in $\mathbf{s}^*$ are arbitrary in the sets $\mathcal{K}$ and $\mathcal{M}$, respectively. It is concluded that if $\alpha^*_{km} = 0$, $\mathbf{u}^*$ admits $\tilde{\mathbf{w}}_{km} = \mathbf{0}$ to generate an optimal solution, which completes the proof.

Appendix B: Proof of Theorem 1

Let us define $\mathbf{u} = (\mathbf{s}, \boldsymbol{\mu}) \in \mathcal{U} \triangleq \{(\mathbf{w}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\mu}) | \mathbf{w}_{km} \in \ker(\mathbf{h}^{\mathtt{d}}_{km})\} \subseteq \mathcal{F}$, $\forall k \in \mathcal{K}$, $m \in \mathcal{M}$, where $\mathbf{s}$ represents the triple $(\mathbf{w}, \mathbf{p}, \boldsymbol{\alpha})$ as part of quadruple $\mathbf{u}$. To prove Theorem 1, we need to show two states: (i) $F(\mathbf{u}|\mathbf{w}_{km} = \mathbf{0}) \geq F(\mathbf{u}|\mathbf{w}_{km} \neq \mathbf{0})$, $\forall \alpha_{km} \in \{0, 1\}$; and (ii) $F(\mathbf{u}|\mathbf{w}_{km} = \mathbf{0} \ \& \ \alpha_{km} = 0) \geq F(\mathbf{u}|\mathbf{w}_{km} \neq \mathbf{0} \ \& \ \alpha_{km} = 1)$. For the first state, we denote $\mathbf{u}_0 \in \{(\mathbf{s}_0, \boldsymbol{\mu}) \in \mathcal{U} | \mathbf{w}_{km} = \mathbf{0}\}$, and $\mathbf{u}_1 \in \{(\mathbf{s}_1, \boldsymbol{\mu}) \in \mathcal{U} | \mathbf{w}_{km} \neq \mathbf{0}\}$ and consider DL SINRs in (2) with respect to $\mathbf{s}_0$ and $\mathbf{s}_1$. For $\mathbf{w}_{km} \in \ker(\mathbf{h}^{\mathtt{d}}_{km})$, it follows that $|\mathbf{h}^{\mathtt{d}}_{km}\mathbf{w}_{km}|^2 = 0$, and thus, $\gamma^{\mathtt{d}}_k(\mathbf{s}_0) = \gamma^{\mathtt{d}}_k(\mathbf{s}_1)$. On the other hand, $\gamma^{\mathtt{d}}_{k'}(\mathbf{s}_0) \geq \gamma^{\mathtt{d}}_{k'}(\mathbf{s}_1)$, $\forall k' \in \mathcal{K} \setminus \{k\}$ implies that the component $|\mathbf{h}^{\mathtt{d}}_{k'm}\mathbf{w}_{km}|^2$ in the denominator of SINR for $\mathtt{U}^{\mathtt{d}}_{k'}$ is equal to or greater than zero, where the equality holds if $\mathbf{w}_{km} = \mathbf{0}$. In addition, we have $F_{\mathtt{SE}}(\mathbf{s}_0) \geq F_{\mathtt{SE}}(\mathbf{s}_1)$ due to $\gamma^{\mathtt{u}}_\ell(\mathbf{s}_0) \geq \gamma^{\mathtt{u}}_\ell(\mathbf{s}_1), \forall \ell \in \mathcal{L}$. Meanwhile, it is true that $\|\mathbf{w}_{km}\|^2 > 0$ for any $\mathbf{w}_{km} \neq \mathbf{0}$, yielding $P_{\mathtt{D}}(\mathbf{u}_0) \leq P_{\mathtt{D}}(\mathbf{u}_1)$ and $P_{\mathtt{T}}(\mathbf{u}_0) \leq P_{\mathtt{T}}(\mathbf{u}_1)$. That is to say $F_{\mathtt{EE}}(\mathbf{u}_0) \geq F_{\mathtt{EE}}(\mathbf{u}_1)$ as well as $F(\mathbf{u}_0) \geq F(\mathbf{u}_1)$, concluding the first state. The second state is easily proved by following the same steps in Appendix A.

Appendix C: Proof of Theorem 2

The power consumption for data transmission and baseband processing $P_{\mathtt{D}}(\mathcal{V}, f_{\mathtt{spr}}(\mathbf{w}), \boldsymbol{\mu})$ is rewritten as

$$P_{\mathtt{D}}(\mathcal{V}, f_{\mathtt{spr}}(\mathbf{w}), \boldsymbol{\mu}) = \sum\nolimits_{m \in \mathcal{M}} \mu_m \Bigg(\sum\nolimits_{k \in \mathcal{K}} \Big(\frac{\|\mathbf{w}_{km}\|^2}{\nu^{\mathtt{AP}}_m} + L P^{\mathtt{u}}_m + r_{\mathtt{sp}}\big(\mathbf{w}_{km}, \mathbf{h}^{\mathtt{d}}_{km} | \mathbf{w}^{(\kappa)}_k, \mathbf{h}^{\mathtt{d}}_k\big) P^{\mathtt{d}}_{km} \Big) \Bigg) + B \cdot F_{\mathtt{SE}}(\mathbf{w}, \mathbf{p}, \boldsymbol{\alpha}) \cdot P^{\mathtt{bh}} + \sum\nolimits_{\ell \in \mathcal{L}} \frac{p_\ell}{\nu^{\mathtt{u}}_\ell}. \quad \text{(C.1)}$$

It can be foreseen that if $\mu_m = 0$, the signal power of all UEs served by $\mathtt{AP}_m$ becomes zero. In other words, μ_m is also coupled with $\mathbf{w}_{km}, \forall k \in \mathcal{K}$, and $\mathbf{a}_{m\ell}, \forall \ell \in \mathcal{L}$. The first term in (C.1) is associated with the DL and UL power allocation, and thus, μ_m must be strictly updated with respect to the DL and UL signal power, showing (17).

Appendix D: Proof of Theorem 3

The proof is done by showing the fact that problems (27) and (28) share the same optimal objective and solution set. From the introduction of soft SINRs $\boldsymbol{\lambda}$, it is straightforward

to prove that constraints (28c) and (28d) must be active (i.e., holding with equalities) at optimum. As a result, constraints (27b) and (27c) can be converted to linear constraints (28e) and (28f), respectively. In addition, we can decompose $\bar{F}_{\mathtt{SE}}\big(\mathbf{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}},\mathbf{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}\big)$ as $\bar{F}_{\mathtt{SE}}\big(\mathbf{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}},\mathbf{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}\big) = R_{\Sigma}(\mathbf{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}})+R_{\Sigma}(\mathbf{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}})$, where

$$\begin{aligned} R_{\Sigma}(\mathbf{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}}) &\geq R_{\Sigma}(\boldsymbol{\lambda}_{\mathtt{d}}) = \sum\nolimits_{k\in\mathcal{K}} \ln(1+\lambda_k^{\mathtt{d}}) \\ &= \ln\Big(\prod\nolimits_{k\in\mathcal{K}}(1+\lambda_k^{\mathtt{d}})\Big) = \ln|\mathbf{I}+\mathbf{\Lambda}_{\mathtt{d}}|, \qquad \text{(D.1a)} \\ R_{\Sigma}(\mathbf{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}) &\geq R_{\Sigma}(\boldsymbol{\lambda}_{\mathtt{u}}) = \sum\nolimits_{\ell\in\mathcal{L}} \ln(1+\lambda_\ell^{\mathtt{u}}) \\ &= \ln\Big(\prod\nolimits_{\ell\in\mathcal{L}}(1+\lambda_\ell^{\mathtt{u}})\Big) = \ln|\mathbf{I}+\mathbf{\Lambda}_{\mathtt{u}}|. \qquad \text{(D.1b)} \end{aligned}$$

Then, we have

$$\begin{aligned} \bar{F}_{\mathtt{SE}}\big(\mathbf{\Gamma}_{\mathtt{d}}^{\mathtt{ZF}},\mathbf{\Gamma}_{\mathtt{u}}^{\mathtt{ZF}}\big) &\geq \ln|\mathbf{I}+\mathbf{\Lambda}_{\mathtt{d}}| + \ln|\mathbf{I}+\mathbf{\Lambda}_{\mathtt{u}}| \\ &:= \tilde{F}_{\mathtt{SE}}\big(\mathbf{\Lambda}_{\mathtt{d}},\mathbf{\Lambda}_{\mathtt{u}}\big). \qquad \text{(D.2)} \end{aligned}$$

Equalities in (28c) and (28d) also lead to an equality of (D.2). In the same spirit, we can further show that constraint (28b) is also active at optimum, which completes the proof.

Appendix E: Proof of Procedure 1

Suppose that a projection matrix $\mathbf{P} = \mathbf{I} - (\tilde{\mathbf{G}}^{\mathtt{AA}})^H\big(\tilde{\mathbf{G}}^{\mathtt{AA}}(\tilde{\mathbf{G}}^{\mathtt{AA}})^H\big)^{-1}\tilde{\mathbf{G}}^{\mathtt{AA}}$ is applied to cancel the IAI and RSI. Accordingly, the effects of MUI, IAI and RSI can be ignored by considering $\bar{\mathbf{W}} = \mathbf{P}\mathbf{Q}^H\tilde{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}}$ as a precoder matrix.

- **MUI cancellation**: it follows that

$$\mathbf{H}^{\mathtt{d}}\bar{\mathbf{W}} \overset{[a]}{=} \mathbf{T}\mathbf{Q}\mathbf{Q}^H\tilde{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}} \overset{[b]}{=} \bar{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}},$$

  where $[a]$ is obtained via Step 4, while $[b]$ comes from the structure of $\tilde{\mathbf{T}}$ in Step 5. Clearly, MUI is completely removed, since both $\bar{\mathbf{T}}$ and $\mathbf{D}^{\mathtt{d}}$ are diagonal matrices.
- **IAI and RSI cancellation**: the effective IAI and RSI are generally expressed by $\tilde{\mathbf{G}}^{\mathtt{AA}}$. We have

$$\tilde{\mathbf{G}}^{\mathtt{AA}}\bar{\mathbf{W}} = \tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{P}\mathbf{Q}^H\tilde{\mathbf{T}}(\mathbf{D}^{\mathtt{d}})^{\frac{1}{2}} = \mathbf{0}.$$

  due to $\tilde{\mathbf{G}}^{\mathtt{AA}}\mathbf{P} = \mathbf{0}$. However, $\tilde{\mathbf{G}}^{\mathtt{AA}}$ is a concatenation of IAI and RSI matrices, leading to a full-rank matrix in most cases. In other words, $\mathbf{P}$ must be forced to $\mathbf{0}$, and thus, should not be joined into the precoder matrix $\bar{\mathbf{W}}$. To overcome this issue, we exploit the PCA-based method to depress the IAI and RSI in the rest of this proof.

To derive matrix $\mathbf{P}$, we consider a low-rank approximation via the PCA method as in Steps 1-3. From (38) and (39), the low-rank approximation of $\tilde{\mathbf{G}}^{\mathtt{AA}}$ can be derived from the $\bar{N}$-top eigenvalues as

$$(\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}})^H\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}} = \bar{\mathbf{U}}\bar{\mathbf{E}}^{\frac{1}{2}}\bar{\mathbf{E}}^{\frac{1}{2}}\bar{\mathbf{U}}^H, \qquad \text{(E.1)}$$

where $\bar{\mathbf{U}} \in \mathbb{C}^{N\times\bar{N}}$ involves the first $\bar{N}$ columns of $\mathbf{U}$, and the diagonal matrix $\bar{\mathbf{E}} \in \mathbb{C}^{\bar{N}\times\bar{N}}$ has the main diagonal with the $\bar{N}$-top eigenvalues of $\mathbf{E}$ in (38). We note that $\bar{\mathbf{U}}^H\bar{\mathbf{U}} = \mathbf{I}$, but $\bar{\mathbf{U}}\bar{\mathbf{U}}^H \neq \mathbf{I}$. By treating $\tilde{\mathbf{G}}^{\mathtt{AA}}$ as $\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}} = \bar{\mathbf{E}}^{\frac{1}{2}}\bar{\mathbf{U}}^H$, the projection matrix $\mathbf{P}$ can be calculated as

$$\begin{aligned} \mathbf{P} &= \mathbf{I} - (\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}})^H\big(\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}}(\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}})^H\big)^{-1}\tilde{\mathbf{G}}_{\bar{N}}^{\mathtt{AA}} \\ &= \mathbf{I} - \bar{\mathbf{U}}\bar{\mathbf{E}}^{\frac{1}{2}}(\bar{\mathbf{E}}^{\frac{1}{2}}\bar{\mathbf{U}}^H\bar{\mathbf{U}}\bar{\mathbf{E}}^{\frac{1}{2}})^{-1}\bar{\mathbf{E}}^{\frac{1}{2}}\bar{\mathbf{U}}^H = \mathbf{I} - \bar{\mathbf{U}}\bar{\mathbf{U}}^H, \qquad \text{(E.2)} \end{aligned}$$

showing Step 3.

Appendix F: Proof of Theorem 4

From $\bar{\mathbf{\Xi}} = \mathbf{\Xi}\mathbf{\Upsilon}$, it is clear that $\bar{\boldsymbol{\varphi}}_j = \mathbf{\Xi}\boldsymbol{\upsilon}_j$ with $\boldsymbol{\upsilon}_j$ being the j-th column of $\mathbf{\Upsilon}$. Therefore, ε_{mj} can be expressed as a function of assignment variables, i.e.,

$$\begin{aligned} \frac{N_m\varepsilon_{mj}}{\beta_{mj}} &= N_m\Big(1 - \frac{\tau p_j^{\mathtt{tr}}\beta_{mj}}{\sum_{j'\in\mathcal{T}_{\mathtt{u}}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}|\boldsymbol{\upsilon}_j^H\mathbf{\Xi}^H\mathbf{\Xi}\boldsymbol{\upsilon}_{j'}|^2 + \sigma_{\mathtt{AP}}^2}\Big) \\ &= N_m\Big(1 - \frac{\tau p_j^{\mathtt{tr}}\beta_{mj}}{\sum_{j'\in\mathcal{T}_{\mathtt{u}}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}|\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'}|^2 + \sigma_{\mathtt{AP}}^2}\Big) \\ &\overset{[a]}{=} N_m\frac{\sum_{j'\in\mathcal{T}_{\mathtt{u}}\setminus\{j\}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} + \sigma_{\mathtt{AP}}^2}{\tau p_j^{\mathtt{tr}}\beta_{mj} + \sum_{j'\in\mathcal{T}_{\mathtt{u}}\setminus\{j\}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} + \sigma_{\mathtt{AP}}^2}, \end{aligned} \qquad \text{(F.1)}$$

where $[a]$ comes from the fact that $\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} \in \{0,1\}$, $\forall j,j' \in \mathcal{T}_{\mathtt{u}}$, and $\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} = 1$ when $j = j'$. In addition, for arbitrary values of x, y and z, such that $0 \leq z < x < y$, it is true that $\frac{x-z}{y-z} \leq \frac{x}{y} \leq \frac{x+z}{y+z}$. Upon setting $x = \sum_{j'\in\mathcal{T}_{\mathtt{u}}\setminus\{j\}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} + \sigma_{\mathtt{AP}}^2$ and $y = x + \tau p_j^{\mathtt{tr}}\beta_{mj}$, z is the amount of disparity in x when $\boldsymbol{\upsilon}_j$ changes. It implies that when $\boldsymbol{\upsilon}_j$ changes, x and y vary by the same amount of z. Consequently, we replace x/y with y for the ease of solution derivation, and thus the objective (50a) can be rewritten as

$$\begin{aligned} &\max_{j\in\mathcal{T}_{\mathtt{u}}}\sum\nolimits_{m\in\mathcal{M}} N_m\sum\nolimits_{j'\in\mathcal{T}_{\mathtt{u}}}\tau p_{j'}^{\mathtt{tr}}\beta_{mj'}\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'} + \sigma_{\mathtt{AP}}^2 \\ &= \max_{j\in\mathcal{T}_{\mathtt{u}}}\sum\nolimits_{j'\in\mathcal{T}_{\mathtt{u}}}\boldsymbol{\upsilon}_j^H\boldsymbol{\upsilon}_{j'}\sum\nolimits_{m\in\mathcal{M}} N_m\tau p_{j'}^{\mathtt{tr}}\beta_{mj'} + \sigma_{\mathtt{AP}}^2. \qquad \text{(F.2)} \end{aligned}$$

Since $\sigma_{\mathtt{AP}}^2$ in (F.2) is the constant, we can arrive at a tractable optimization problem (51).

References

[1] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, S. K. Sharma, S. Chatzinotas, B. Ottersten, and O.-S. Shin, "A novel heap-based pilot assignment for full duplex cell-free massive MIMO with zero-forcing," in *IEEE Inter. Conf. Commun. (ICC)*, Dublin, Ireland, June 2020, pp. 1–6, to appear.

[2] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*, Mar. 2017.

[3] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G Mobile and Wireless Communications Technology*. Cambridge University Press, 2016.

[4] A. Yadav and O. A. Dobre, "All technologies work together for good: A glance at future mobile networks," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 10–16, Aug. 2018.

[5] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Multiple antenna technologies for beyond 5G," Sept. 2019. [Online]. Available: https://arxiv.org/abs/1910.00092

[6] S. Chatzinotas, M. A. Imran, and R. Hoshyar, "On the multicell processing capacity of the cellular MIMO uplink channel in correlated rayleigh fading environment," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3704–3715, July 2009.

[7] S. Buzzi *et al.*, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Select. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[8] A. Yadav, G. I. Tsiropoulos, and O. A. Dobre, "Full-duplex communications: Performance in ultradense mm-wave small-cell wireless networks," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 40–47, June 2018.

[9] S. K. Sharma *et al.*, "Dynamic spectrum sharing in 5G wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Commun. Surveys Tutor.*, vol. 20, no. 1, pp. 674–707, Firstquarter 2018.

[10] A. Sabharwal *et al.*, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Select. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Feb. 2014.

[11] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: Will doubling interference prevent doubling capacity?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.

[12] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM Computer Commun. Review*, 2013, pp. 375–386.

[13] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. 11th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, Seattle, WA, USA, 2014, pp. 369–372.

[14] A. Yadav, O. A. Dobre, and N. Ansari, "Energy and traffic aware full duplex communications for 5G systems," *IEEE Access*, vol. 5, pp. 11 278–11 290, May 2017.

[15] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896–4910, Sept. 2014.

[16] V.-D. Nguyen *et al.*, "Spectral efficiency of full-duplex multiuser system: Beamforming design, user grouping, and time allocation," *IEEE Access*, vol. 5, pp. 5785–5797, Mar. 2017.

[17] H. V. Nguyen *et al.*, "Joint antenna array mode selection and user assignment for full-duplex MU-MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 2946–2963, June 2019.

[18] H. V. Nguyen *et al.*, "Joint power control and user association for NOMA-based full-duplex systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8037–8055, Nov. 2019.

[19] V.-D. Nguyen, H. V. Nguyen, O. A. Dobre, and O.-S. Shin, "A new design paradigm for secure full-duplex multiuser systems," *IEEE J. Select. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, July 2018.

[20] V.-D. Nguyen *et al.*, "Spectral and energy efficiencies in full-duplex wireless information and power transfer," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2220–2233, May 2017.

[21] H. H. M. Tam, H. D. Tuan, and D. T. Ngo, "Successive convex quadratic programming for quality-of-service management in full-duplex MU-MIMO multicell networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2340–2353, June 2016.

[22] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, Apr. 2017.

[23] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[24] E. Nayebi *et al.*, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, July 2017.

[25] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "How much do downlink pilots improve cell-free massive MIMO?" in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.

[26] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max–min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.

[27] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.

[28] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[29] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[30] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[31] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[32] B. K. Chalise *et al.*, "Beamforming optimization for full-duplex wireless-powered MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3750–3764, Sept. 2017.

[33] T. T. Vu *et al.*, "Full-duplex cell-free massive MIMO," in *Proc. IEEE Inter. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[34] D. Wang, M. Wang, P. Zhu, J. Li, J. Wang, and X. You, "Performance of network-assisted full-duplex for cell-free massive MIMO," May 2019. [Online]. Available: http://arxiv.org/abs/1905.11107

[35] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[36] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, July-Aug. 1978.

[37] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[38] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[39] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[40] S. Tombaz *et al.*, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.

[41] E. Björnson *et al.*, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, June 2015.

[42] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization.* Philadelphia: MPS-SIAM Series on Optimi., SIAM, 2001.

[43] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge Univ. Press, UK, 2005.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[45] T. H. Cormen *et al.*, *Introduction to Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, 2001.

[46] D. Korpi, M. Heino, C. Icheln, K. Haneda, and M. Valkama, "Compact inband full-duplex relays with beyond 100 dB self-interference suppression: Enabling techniques and field measurements," *IEEE Trans. Antennas Propagation*, vol. 65, no. 2, pp. 960–965, Feb. 2017.

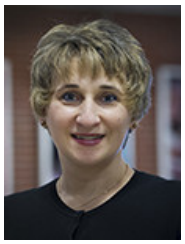
[47] Ao Tang, JiXian Sun, and Ke Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE Veh. Tech. Conf. (VTC Spring).*, May 2001, pp. 333–336.

[48] H. M. Kim, H. V. Nguyen, G.-M. Kang, Y. Shin, and O.-S. Shin, "Device-to-device communications underlaying an uplink SCMA system," *IEEE Access*, vol. 7, pp. 21 756–21 768, Feb. 2019.

Hieu V. Nguyen (S'16) received the B.S. degree in electronics and telecommunications from Danang University of Science and Technology, Danang, Vietnam, in 2011, and the M.S. degree in electronic engineering from Soongsil University, Seoul, Korea, in 2016. Since 2016, he has been pursuing the Ph.D. degree in the School of Electronic Engineering, Soongsil University, Seoul, Korea. From 2011 to 2013, he was with Danang University of Science and Technology as an Assistant Researcher and a Lecturer. Since 2014, he has been with Wireless Communications Laboratory at Soongsil University. His research interest is in wireless communications, with particular focus on optimization techniques and machine learning for wireless communications, UAV/drones communications, device-to-device communications, full-duplex radios, energy harvesting, and cognitive radio systems.

Van-Dinh Nguyen (S'14-M'19) received the B.E. degree in electrical engineering from Ho Chi Minh City University of Technology, Vietnam, in 2012 and the M.E. and Ph.D. degrees in electronic engineering from Soongsil University, Seoul, South Korea, in 2015 and 2018, respectively. He is currently a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. He was a Postdoc Researcher and a Lecturer with Soongsil University, a Postdoctoral Visiting Scholar with University of Technology Sydney, AUS (July-August 2018) and a Ph.D. Visiting Scholar with Queen's University Belfast, U.K. (June-July 2015 and August 2016). His current research activity is focused on the mathematical modelling of 5G cellular networks and machine learning for wireless communications.

Dr. Nguyen received several best conference paper awards, IEEE Transaction on Communications Exemplary Reviewer 2018 and IEEE GLOBECOM Student Travel Grant Award 2017. He has authored or co-authored in some 40 papers published in international journals and conference proceedings. He has served as a reviewer for many top-tier international journals on wireless communications, and has also been a Technical Programme CommitteeMember for several flag-ship international conferences in the related fields. He is an Editor for the IEEE Open Journal of the Communications Society and IEEE Communications Letters.



Octavia A. Dobre (M'05-SM'07-F'20) received the Dipl. Ing. and Ph.D. degrees from Politehnica University of Bucharest (formerly Polytechnic Institute of Bucharest), Romania, in 1991 and 2000, respectively. Between 2002 and 2005, she was with New Jersey Institute of Technology, USA and Politehnica University of Bucharest. In 2005, she joined Memorial University, Canada, where she is currently Professor and Research Chair. She was a Visiting Professor with Massachusetts Institute of Technology, USA and Université de Bretagne Occidentale, France.

Her research interests include enabling technologies for beyond 5G, blind signal identification and parameter estimation techniques, as well as optical and underwater communications. She authored and co-authored over 250 refereed papers in these areas.

Dr. Dobre serves as the Editor-in-Chief (EiC) of the IEEE Open Journal of the Communications Society and Editor of the IEEE Communications Surveys and Tutorials and IEEE Vehicular Technology Magazine. She was the EiC of the IEEE Communications Letters, as well as Senior Editor, Editor, and Guest Editor for various prestigious journals and magazines. Dr. Dobre was the General Chair, Technical Program Co-Chair, Tutorial Co-Chair, and Technical Co-Chair of symposia at numerous conferences.

Dr. Dobre was a Royal Society Scholar and a Fulbright Scholar. She obtained Best Paper Awards at various conferences, including IEEE ICC, IEEE Globecom and IEEE WCNC. Dr. Dobre is a Distinguished Lecturer of the IEEE Communications Society. Dr. Dobre is Fellow of the Engineering Institute of Canada.

She is a member-at-large of the Board of Governors of the IEEE Communications Society, has served with the Administrative Committee of the IEEE Instrumentation and Measurement Society, as well as with many other committees in these professional societies.

Shree Krishna Sharma (S'12-M'15-SM'18) is currently Research Scientist at the SnT, University of Luxembourg. Prior to this, he worked as a Postdoctoral Fellow at the Western University, Canada, and as a Research Associate at the SnT being involved in different European, national and ESA projects after receiving his PhD degree in Wireless Communications from the University of Luxembourg in 2014. His current research interests include 5G and beyond wireless, Internet of Things, machine learning, edge computing and optimization of distributed communications, computing and caching resources.

He has published about 100 technical papers in scholarly journals, international conferences and book chapters, and has over 2000 google scholar citations. He is a Senior Member of IEEE and is the recipient of several prestigious awards including "2018 EURASIP JWCN Best Paper Award", "CROWNCOM 2015 Best Paper Award" and "FNR Award for Outstanding PhD Thesis 2015", and the co-recipient of "FNR Award for Outstanding Scientific Publication 2019". He has been serving as a Reviewer for several international journals and conferences; as a TPC member for a number of international conferences including IEEE ICC, IEEE GLOBECOM, IEEE PIMRC, IEEE VTC and IEEE ISWCS; and an Associate Editor for IEEE Access journal. He co-organized a special session in IEEE PIMRC 2017, a workshop in IEEE SECON 2019, worked as a Track co-chair for IEEE VTC-fall 2018 conference, and published an IET book on "Satellite Communications in the 5G Era" as a lead editor.

Symeon Chatzinotas is currently Full Professor/Chief Scientist I and Co-Head of the SIGCOM Research Group at SnT, University of Luxembourg. In the past, he has been a Visiting Professor at the University of Parma, Italy and he was involved in numerous Research and Development projects for the National Center for Scientific Research Demokritos, the Center of Research and Technology Hellas and the Center of Communication Systems Research, University of Surrey. He received the M.Eng. degree in telecommunications from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, Surrey, U.K., in 2006 and 2009, respectively. He was a co-recipient of the 2014 IEEE Distinguished Contributions to Satellite Communications Award, the CROWNCOM 2015 Best Paper Award and the 2018 EURASIC JWCN Best Paper Award. He has (co-)authored more than 400 technical papers in refereed international journals, conferences and scientific books. He is currently in the editorial board of the IEEE Open Journal of Vehicular Technology and the International Journal of Satellite Communications and Networking.

Björn Ottersten (S'87–M'89–SM'99–F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Ottersten has been Head of the Department for Signals, Sensors, and Systems, KTH, and Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg.

He is a recipient of the IEEE Signal Processing Society Technical Achievement Award and the European Research Council advanced research grant twice. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and 8 IEEE conference papers best paper awards. He has been a board member of IEEE Signal Processing Society, the Swedish Research Council and currently serves of the boards of EURASIP and the Swedish Foundation for Strategic Research. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the Editorial Board of the IEEE Signal Processing Magazine. He is currently a member of the editorial boards of IEEE Open Journal of Signal Processing, EURASIP Signal Processing Journal, EURASIP Journal of Advances Signal Processing and Foundations and Trends of Signal Processing. He is a fellow of EURASIP.

Oh-Soon Shin (S'00-M'10) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 1998, 2000, and 2004, respectively. From 2004 to 2005, he was with the Division of Engineering and Applied Sciences, Harvard University, MA, USA, as a Post-Doctoral Fellow. From 2006 to 2007, he was a Senior Engineer with Samsung Electronics, Suwon, Korea. In September 2007, he joined the School of Electronic Engineering, Soongsil University, Seoul, Korea, where he is currently a Professor. His research interests include communication theory, wireless communication systems, and signal processing for communication.